Joint modelling of events and repeated observations : an application to the progression of Amyotrophic Lateral Sclerosis

Juliette Ortholand

Sorbonne Université

Doctoral Thesis

Joint modelling of events and repeated observations: an application to the progression of Amyotrophic Lateral Sclerosis

---

Author:
Juliette Ortholand

Supervisors:
Sophie Tezenas du Montcel
Stanley Durrleman

Referees:
Agathe Guilloux
Hélène Jacqmin-Gadda
Florence Forbes
Olivier Bouaziz

A thesis submitted in fulfilment of the requirements for the degree of PhD in Sciences and Technologies of Information and Communication

in the

ARAMIS Lab
Sorbonne Université, Institut du Cerveau - Paris Brain Institute (ICM), CNRS, Inria, Inserm, AP-HP Hôpital de la Pitié Salpêtrière

June 6, 2024

All models are wrong, but some are useful. (Box, 1979)

Abstract

Progression heterogeneity in chronic diseases such as Amyotrophic Lateral Sclerosis (ALS) is a significant obstacle to developing effective treatments. Leveraging the growing wealth of large databases through modelling can help better understanding it. However, the data collected only offer access to partial trajectories, that need to be realigned to reconstruct a comprehensive disease progression.

To address this challenge, data-driven progression models like the longitudinal Spatiotemporal model were developed. Its main interest is its ability to synchronise patients onto a common disease timeline (temporal aspect) thanks to a latent disease age, while also capturing the remaining variability through parameters that account for outcome ordering (spatial aspect). However, this model was primarily designed for longitudinal data, overlooking crucial outcomes in ALS such as time to death or initiation of life support, like Non-Invasive Ventilation (NIV). Conversely, existing joint models offer the advantage of simultaneously handling longitudinal and survival data. However, they do not realign trajectories, which compromises their temporal resolution.

This thesis aimed to expand the Spatiotemporal model into a Joint Spatiotemporal model, enabling, for ALS research, the examination of survival data alongside longitudinal data.

First, we applied the Spatiotemporal model to explore how the interaction between sex and onset site (spinal or bulbar) impacts the progression of ALS patients. We selected 1,438 patients from the PRO-ACT database. We demonstrated a significant influence of both sex and onset site on six longitudinal outcomes monitoring the functional and respiratory decline in addition to Body Mass Index. However, this study did not incorporate survival analysis, despite its paramount importance in ALS, due to limitations inherent to the Spatiotemporal model.

To address this gap, we associated the Spatiotemporal model with a survival model that estimates a Weibull survival model from its latent disease age, creating a univariate Joint Temporal model. After model validation, we benchmarked our model with a state-of-the-art joint model on PRO-ACT data. Our model exhibited significantly superior performance in terms of absolute bias and mean cumulative AUC for right-censored events. This demonstrated the efficacy of our approach in the context of ALS compared to existing joint models. However, modelling several longitudinal outcomes requires a multivariate approach. Life support initiation that might be censored by death needs to be also considered.

We thus extended the Joint Temporal model, into a multivariate Joint Spatiotemporal model with competing risks to analyse NIV initiation. This involved coupling the multivariate Spatiotemporal model with a cause-specific Weibull survival model from the latent disease age. We incorporated spatial parameters with a Cox proportional effect on the hazard. After validation, we benchmarked our model with a state-of-the-art joint model on PRO-ACT data and analysed sex and onset site interaction in complement to the first study. The Joint Spatiotemporal model achieved similar performance to the state-of-the-art model while capturing an underlying shared latent process, the latent disease age, whereas the state-of-the-art models the impact of longitudinal outcomes on survival.

To enhance the reproducibility and facilitate the reuse of these models, the proposed models were implemented in the open-source software Leaspy.

In conclusion, this thesis introduces the first data-driven progression model combining longitudinal and survival modelling. We demonstrated its relevance to understand the occurrence of critical events in ALS. This work paves the way for further extension to analyse recurrent events, among other potential applications in causal inference.

Résumé

L'hétérogénéité des progressions des maladies chroniques, comme la Sclérose Latérale Amyotrophique (SLA), constitue un obstacle au développement de traitements. L'utilisation croissante de bases de données, couplée à la modélisation de ces maladies, contribue à une meilleure compréhension de ce phénomène. Cependant, les données collectées ne permettent de décrire que des trajectoires partielles qui doivent être réalignées pour reconstruire une progression complète de la maladie.

Le modèle spatiotemporel est l'un des modèles développés pour traiter cette question. Son principal intérêt est sa capacité à réaligner les progressions des patients à la fois en terme de chronologie (aspect temporel), grâce à un âge latent, et en terme d'ordre de progression des données longitudinales (aspect spatial). Cependant, ce modèle n'a pas été conçu pour la modélisation d'évènements, cruciaux dans la SLA : survie ou introduction de support de vie, comme la ventilation non invasive (VNI). A l'inverse, les modèles conjoints existants offrent l'avantage de traiter simultanément des données longitudinales et de survie. Cependant, ils ne réalignent pas les trajectoires, ce qui compromet leur résolution temporelle.

Cette thèse visait à étendre le modèle spatiotemporel en un modèle spatiotemporel conjoint, permettant l'examen des données longitudinales et de survie dans la SLA. Nous avons d'abord appliqué le modèle spatiotemporel pour étudier l'impact de l'interaction entre le sexe et le site de début (spinal ou bulbaire) sur la progression de la SLA. En utilisant la base de données PRO-ACT, nous avons démontré une influence significative à la fois du sexe et du site de début sur six résultats longitudinaux surveillant le déclin fonctionnel et respiratoire ainsi que l'indice de masse corporelle. Cependant, cette étude n'a pas incorporé d'analyse de survie, malgré son importance dans la SLA, en raison des limitations du modèle.

Pour combler cette lacune, nous avons associé au modèle spatiotemporel un modèle de survie qui estime une distribution de Weibull de l'évènement (le décès ici) à partir de l'âge latent, créant ainsi un modèle temporel univarié conjoint. Après validation du modèle, nous l'avons comparé à un modèle conjoint de l'état de l'art sur les données PRO-ACT. Notre modèle a obtenu des performances significativement supérieures en termes de biais absolu et d'AUC moyenne pour les événements censurés à droite. Cela a démontré l'efficacité de notre approche dans le contexte de la SLA. Cependant, la modélisation de plusieurs résultats longitudinaux nécessite une approche multivariée et l'initiation de support de vie, qui peut être censurée par le décès, doit aussi être prise en compte.

Nous avons donc étendu le modèle temporel conjoint en un modèle spatiotemporel conjoint avec des risques concurrents pour analyser l'initiation de la VNI. Cela impliquait de coupler le modèle spatiotemporel multivarié avec un modèle de survie pour risques concurrents estimé à partir de l'âge latent. Après validation, nous avons comparé notre modèle à un modèle conjoint de l'état de l'art sur les données PRO-ACT et avons analysé l'interaction entre le sexe et le site de début en complément de la première étude. Le modèle spatiotemporel conjoint a atteint des performances similaires au modèle de l'état de l'art tout en capturant un processus latent partagé, l'âge latent, tandis que le modèles de l'état de l'art examinent l'impact des données longitudinales sur la survie.

Pour faciliter la reproductibilité et la réutilisation de ces modèles, ils ont été implémentés dans le logiciel open source Leaspy.

Cette thèse introduit le premier modèle permettant de réaligner des trajectoires partielles en combinant la modélisation de données longitudinales et de survie. Nous avons démontré sa pertinence pour comprendre l'occurrence d'événements critiques dans la SLA.

Acknowledgements

Ce travail n'aurait pu aboutir sans le soutien et l'accompagnement de nombreuses personnes, auxquelles j'aimerais exprimer ici toute ma reconnaissance.

Tout d'abord, un grand merci à Hélène Jacqmin-Gadda et Agathe Guilloux pour avoir accepté de relire ce manuscrit de thèse. Leurs remarques ont été précieuses pour prendre du recul sur le travail effectué. Merci également à Olivier Bouaziz et Florence Forbes, membres du jury, dont les questions m'ont permis d'approfondir ma réflexion.

Je souhaite remercier profondément mes directeur-ices de thèse, pour leur encadrement aussi bien scientifique qu'humain. Stanley Durrleman, pour sa confiance et pour m'avoir transmis l'art de raconter des histoires en recherche. Sophie Tezenas du Montcel, pour avoir rejoint l'aventure en cours de route et m'avoir initiée à la rigueur statistique.

Un merci sincère à Pierre-François Pradat, pour son aide précieuse dans la compréhension de la sclérose latérale amyotrophique (SLA) et des questions scientifiques qui la traversent. Merci à Cécile Proust-Lima pour nos échanges, notamment lors de mes comités de suivi, et dont les travaux, en parti communs avec ceux d'Hélène Jacqumin-Gadda, ont été une source d'inspiration pour moi.

Le laboratoire scientifique ARAMIS dans lequel j'ai eu la chance extraordinaire de réaliser ma thèse, m'a aussi énormément apporté d'un point de vue scientifique et humain. Ma reconnaissance va d'abord aux membres de l'équipe Longitudinal et du groupe de travail pour le logiciel Leaspy. Leur accompagnement dans mes questionnements scientifiques, lors de conférences ou de Merge Request, m'a permis de repousser mes limites et de grandir scientifiquement. Mes remerciements vont aussi aux membres permanents, notamment Olivier Colliot, Ninon Burgos et Baptiste Couvy-Duchesne, dont l'excellence académique fut à la fois un appui et une source d'inspiration au quotidien. S'ajoute plus largement l'ensemble des membres d'ARAMIS, ceux avec qui j'ai partagé mes débuts dans la salle Azote et ceux avec qui j'ai continué dans l'annexe de l'open-space. Merci aux anciens pour leurs conseils éclairés, et à ceux avec qui des discussions stimulantes ont été partagées, à la cantine, lors des afterworks ou à vélo. Dans l'ensemble, l'expertise technique des membres du laboratoire, ainsi que leur bonne humeur, ont contribué à rendre ces années de thèse d'autant plus appréciables.

Un grand merci également à mes proches, à mes amis thésards – désormais pour la plupart docteurs – pour leur soutien et conseils avisés. Merci à mes amis d'école, notamment à la "famille Fondue", pour avoir rempli ces dernières années d'aventures et de moments chaleureux, ainsi qu'aux "Trois Grâces" pour leurs discussions sans fin et leur soutien infaillible. Un remerciement à celles et ceux qui m'ont empêché de devenir trop parisienne : les amis de Saint-Paul, de Varces, et plus particulièrement à celles qui vivent libres, pour leur précieuse amitié.

Enfin, une gratitude éternelle à ma famille. A mon parrain et ma marraine pour m'avoir inspiré sur le chemin de la thèse et à mes grands-parents qui m'ont encouragé toutes ces années. Merci à ma soeur, mon frère et mes parents pour leur soutien indéfectible qui m'a donné la force d'entreprendre et d'aller au bout de cette aventure.

Pour conclure, un remerciement tout particulier à Thomas Bonte, pour ses réflexions scientifiques, mais surtout pour tout le reste.

Scientific production

Journal papers

1. M. Dibling, **J. Ortholand**, F. Salachas, A. Hesters, S. Tezenas Du Montcel, *Care pathway heterogeneity in Amyotrophic Lateral Sclerosis: effects of gender, age and onset*, Neuroepidemiology, 2024.

2. **J. Ortholand**, P.-F. Pradat, S. Tezenas Du Montcel, S. Durrleman, *Interaction of sex and onset site on the disease trajectory of ALS*, Journal of Neurology, 2023. [hal-04216957]

3. E. Maheux, I. Koval, **J. Ortholand**, C. Birkenbihl, D. Archetti, V. Bouteloup, S. Epelbaum, C. Dufouil, M. Hofmann-Apitius, S. Durrleman, *Forecasting individual progression trajectories in Alzheimer's disease*, Nature Communications, 2023. [hal-04023781]

Submitted journal papers

1. **J. Ortholand**, S. Durrleman, S. Tezenas Du Montcel, *Joint model with latent disease age: overcoming the need for reference time*, [preprint]

Conferences talks

1. **J. Ortholand**, S. Tezenas Du Montcel, S. Durrleman, *A joint Spatiotemporal multivariate model with competing risks.* IBC 2024 - International Biometric Conference, December 2024, Atlanta, USA.

2. **J. Ortholand**, S. Tezenas Du Montcel, S. Durrleman, *A joint Spatiotemporal multivariate model with competing risks.* ISCB 2024 - 45th Annual Conference of the International Society for Clinical Biostatistics, July 2024, Thessaloniki, Greece.

3. **J. Ortholand**, S. Durrleman, S. Tezenas Du Montcel, *Joint model with latent disease age Overcoming the need for reference time*, Journées de Biostatistique 2023, G.D.R. "Statistique et Santé"; Société française de biométrie; Groupe "Biopharmacie et Santé" de la Société française de statistiques, Nov 2023, Toulouse, France. [hal-04403927]

Conference posters

1. **J. Ortholand**, N. Fournier, P.-E. Poulet, S. Kaisaridi, N. Gensollen, E. Maheux, S. Tezenas du Montcel, S. Durrleman, *Leaspy: LEArning Spatiotemporal patterns in PYthon.* ISCB 2024 - 45th Annual Conference of the International Society for Clinical Biostatistics, July 2024, Thessaloniki, Greece.

2. **J. Ortholand**, S. Durrleman, S. Tezenas Du Montcel. *Longitudinal and survival joint prediction: time reparameterization in amyotrophic lateral sclerosis context.* ISCB 2023 - 44th Annual Conference of the International Society for Clinical Biostatistics, Aug 2023, Milan, Italy. [hal-04216888]

3. **J. Ortholand**, P.-F. Pradat, S. Tezenas Du Montcel, S. Durrleman, *Gender and spinal/bulbar onset interaction on ALS progression.* ISCB 2022 - 43rd Annual Conference of the International Society for Clinical Biostatistics, Aug 2022, Newcastle, United Kingdom. [hal-03782607]

4. **J. Ortholand**,P.-F. Pradat, S. Tezenas Du Montcel, S. Durrleman, *Interaction of gender and onset site in the progression of ALS - A multivariate disease modelling approach.* MNDA - 33rd International Symposium on ALS/MND, Dec 2022, Online, France.

5. **J. Ortholand**,P.-F. Pradat, S. Tezenas Du Montcel, S. Durrleman, *Gender and spinal/bulbar onset interaction on ALS progression – Validation of a multivariate disease course approach*, 8èmes Journées de la Recherche sur la SLA et les Maladies du Neurone Moteur, Oct 2022, Paris, France.

6. **J. Ortholand**, I. Koval, P.-F. Pradat, S. Tezenas Du Montcel, S. Durrleman, *Gender and spinal/bulbar onset interaction on ALS progression.* ENCALS 2022 - European Network to Cure ALS meeting, Jun 2022, Edinburgh, United Kingdom. [hal-03782549]

Contents

List of Figures

List of Tables

List of Abbreviations

AFT	Accelerated Failure Time model
ALS	Amyotrophic Lateral Sclerosis
ALSFRS	ALS Functional Rating Scale
ALSFRSr	revised version of the ALS Functional Rating Scale
ANOVA	ANalysis Of VAriance
AUC	Area Under the Curve
BMI	Body Mass Index
CADASIL	Cerebral Autosomal Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy
CI	Confidence Interval
CIF	Cumulate Incidence Function
C-index	Concordance Index
DAG	Directed Acyclic Graph
ENCALS	European Network for the Cure of ALS
FDA	U.S. Food and Drug Administration
f-MRI	functional Magnetic Resonance Imaging
FT9	Fine'til 9 score
FVC	Forced Vital Capacity
GLMM	Generalised Linear Mixed-effect Model
HR	Hazard Ratio
IBS	Integrated Brier Score
ICA	Independent Component Analysis
JM	Joint Model
Leaspy	LEArning Spatiotemporal patterns in PYthon
LMM	Linear Mixed-effect Model
MCMC-SAEM	Monte-Carlo Markov Chain Stochastic Approximation Expectation-Maximization
MDC	Minimal Detectable Change
MDS-UPDRS	Movement Disorder Society sponsored revision of the Unified Parkinson's Disease Rating Scale
Mitos	ALS Milano-Torino staging
MMSE	Mini-Mental State Examination
NIV	Non Invasive Ventillation
PRO-ACT	Pooled Resource Open-Access ALS Clinical Trials
SD	Standard Deviation
SOC	Standard Of Care
TADPOLE	The Alzheimer's Disease Prediction Of Longitudinal Evolution
WAIC	Watanabe–Akaike Information Criterion

Introduction

The challenge of age-related diseases for public health

In 2023, according to the World Health Organisation, chronic diseases cause 74% of all deaths globally (Organization et al., 2022). Chronic disease can be defined as non-contagious diseases with a long latency period and a prolonged temporal course (McKenna et al., 1998). They encompass diseases such as cardiovascular diseases, cancer, and neurodegenerative diseases. With the ageing population, neurodegenerative disorders are an important burden for the healthcare system (Zahra et al., 2020). Neurodegenerative diseases are caused by the progressive loss of structure or function of neurons, in the process known as neurodegeneration. The three most incident neurodegenerative diseases are Alzheimer's disease (11.08 per 1,000 person-years in Europe (Niu et al., 2017)), Parkinson's disease (11–19 per 100,000 person-years in Europe (Balestrino and Schapira, 2020)) and Amyotrophic Lateral Sclerosis (ALS) (1-2.6 per 100,000 in Europe (Talbott, Malek, and Lacomis, 2016)).

These diseases have different mechanisms and presentations. Alzheimer's disease is linked to the accumulation of tau and amyloïd beta protein in the brain and causes short-term memory loss, which worsens over time. Parkinson's disease is characterised by the loss of dopaminergic neurons which causes tremors as well as stiffness or slowing of movement. ALS is linked to the death of motoneurone, and results in a loss of voluntary motricity and trouble breathing. Age at onset and time of survival also vary quite a lot with ALS being one of the youngest diseases (58–60 years for ALS (Talbott, Malek, and Lacomis, 2016)) and the fastest mortal one (average survival from onset to death is 3–4 years (Talbott, Malek, and Lacomis, 2016)).

Despite evident differences in pathophysiology and symptoms, these diseases also share important similarities. They are long and progressive with suspected onset occurring years before clinical manifestations. For instance, in Alzheimer's disease, the progression of amyloid and neurodegeneration biomarkers starts before the manifestation of the first symptoms (Jack et al., 2010). Similarly in ALS, changes in metabolism start before a significant weight loss and the first motor symptoms (Peter et al., 2017). Neurodegenerative diseases also have heterogeneous progressions with large variations in the disease presentation, speed of progression, and age at onset (Greenland, Williams-Gray, and Barker, 2019; Duara and Barker, 2022; Beghi et al., 2007). The incidence of all these diseases increases with age, making age the most important risk factor in all cases.

There is also no cure for any of these diseases. An ever-increasing research effort to find treatments has resulted in an always-increasing number of clinical trials in this area:

the National Institute on Aging (NIA) is currently supporting 508 active clinical trials on Alzheimer's disease (Aging, 2023). 125 trials on ALS were conducted between 2008 and 2019 (Wong et al., 2021). Only symptomatic treatments exist for Alzheimer's disease (memantine and donepezil (Zahra et al., 2020) and the recently approved lecanemab) and Parkinson's disease (Levodopa and other dopaminergic drugs (Bie et al., 2020)). For ALS, Riluzole enables a significant survival improvement of 3 months on average (Bensimon, Lacomblez, and Meininger, 1994; Miller et al., 2007) and is often associated with life-support clinical procedures, such as gastrostomy to improve feeding, or None-Invasive Ventilation (NIV), to avoid respiratory failures (Bourke et al., 2006; Vianello et al., 2011; Körner et al., 2013). Failure rates in clinical trials are among the highest in this therapeutic area (Cummings, 2018). The reasons are multiple (Yiannopoulou et al., 2019; Petrov et al., 2017; Olanow, Kieburtz, and Schapira, 2008), and among them: the heterogeneity in disease progression (Beghi et al., 2007), the absence of accurate estimates of the disease stage and reliable biomarkers to track disease progression.

Data: a blessing or a curse to understand progression heterogeneity

The development of disease progression models has helped in better understanding this heterogeneity (Locascio and Atri, 2011), which has led to great promise for the design of more powered trials (Maheux et al., 2023; Raket, 2022). To have representative models, important research efforts allowed the creation of a large multimodal longitudinal dataset to study the natural progression of the disease, such as the Alzheimer's disease neuroimaging initiative (ADNI Mueller et al., 2005), the Parkinson's Progression Marker's Initiative (PPMI, Parkinson Progression Marker Initiative, 2011), and NISALS (Müller et al., 2016), PULSE (University Hospital, Lille, 2022), PRO-ACE (Sherman et al., 2019) and PRO-ACT (Atassi et al., 2014) in ALS.

In these studies, patients are followed over several months or years, with the collection of repeated measures of clinical scales or biomarkers at several visits and the recording of events during the follow-up period. These datasets allow the study of disease progression in contrast to cross-sectional studies (Locascio and Atri, 2011). In Alzheimer's and Parkinson's diseases, longitudinal outcomes are collected measuring cognitive (MMSE (Tombaugh and McIntyre, 1992)) or motor (MDS-UPDRS (Goetz et al., 2008)) impairment. In ALS, due to the short survival, and FDA requirement for drug approval (FDA, 2019), death is a major outcome and is often recorded in addition to longitudinal outcomes measuring motor or respiratory decline (ALSFRSr (Rooney et al., 2017), Forced Vital Capacity (Daghlas, Govindarajan, and Pooled Resource Open-Access ALS Clinical Trials Consortium, 2021)). Initiation of life-support clinical procedures is also an important step for the patient's life and is representative of an advanced stage of the disease.

As in many other domains, models face the challenge of the quality of the data collected. Classical challenges of longitudinal data encompass modelling outcomes with low test-retest

reliability (consistency of a measure from one time to another) and collected at irregular timing. The number of patients followed, as well as their duration of follow-up, is also key to model the high heterogeneity of these diseases. Nevertheless, as it was well described by Young et al., 2024, following patients for decades and collecting outcomes (sometimes expensive and invasive) is impossible. Databases result in short-term heterogeneous longitudinal outcomes, observed at different stages of the diseases. The main challenge is thus to realign the partial trajectories to reconstruct a comprehensive disease progression.

There is thus a need to model the heterogeneity of ALS progression, by capturing both longitudinal and survival processes while considering the specificity of the available data.

Contributions

Three types of contributions were made during this thesis. First, methodological contributions were made. We have extended an existing longitudinal model: the Spatiotemporal model. Its strength is its ability to synchronise patients onto a common disease timeline (temporal aspect) thanks to a latent disease age, while also capturing the remaining variability through parameters that account for outcome ordering (spatial aspect). In chapter 3, we have created a univariate Joint Temporal model by associating a Weibull distribution from the latent age to the Spatiotemporal model to also model event with non-informative right censoring (see section 1.3). Then, in chapter 4, we have extended this model into a multivariate Joint cause-specific Spatiotemporal model to handle competing risks (see section 1.3). Both models were validated on simulated data and then benchmarked against reference joint models.

We also applied our models to give better insights into the progression of ALS. To characterise ALS heterogeneity, in chapter 2, we have applied the Longitudinal Spatiotemporal model to a large multicentric ALS dataset, PRO-ACT. In complement to existing cross-sectional and survival analysis, we analysed the interaction between sex and site of onset on the joint progression of six longitudinal outcomes monitoring the functional and respiratory decline in addition to Body Mass Index. Then, thanks to the new Joint cause-specific Spatiotemporal model, in chapter 4, we have complemented the first analysis by taking into account NIV initiation and death.

Finally, software contributions were also made during this thesis. A large refactoring of an existing open-source software named Leaspy for LEArning Spatiotemporal patterns in PYthon was started and the models developed in this thesis were implemented in the new structure, to ease their reuse. All the data processing needed for this thesis was integrated into an internal data processing software for publicly available datasets, to facilitate reproducibility.

Manuscript overview

The manuscript is organized as follows:

- In chapter 1, we reviewed the different models available for longitudinal and survival data associated with the different challenges they tackle, which enabled us to find a well-suited model for longitudinal data for ALS: the Spatiotemporal model.
- In chapter 2, we presented the Spatiotemporal model in more detail and applied it to ALS data to better characterise the heterogeneity of the disease, studying sex and onset site interaction.
- In chapter 3, we extended the Spatiotemporal model in a univariate Joint Temporal model to analyse jointly the functional decline and death. The model was validated on simulated data and benchmarked in both simulated and real data against reference models on PRO-ACT data.
- In chapter 4, we extended the Joint Temporal model into a multivariate Joint cause-specific Spatiotemporal model. After validation, we benchmarked our model with reference models on PRO-ACT data and analysed sex and onset site interaction on NIV initiation and death, in complement to the first study of chapter 2.
- In chapter 5, we detailed our software contributions to both the existing package Leaspy and the internal data processing software.

Chapter 1

Joint models: modelling longitudinal data jointly with events

This chapter is a literature review on longitudinal and survival modelling, focusing on their application in the context of ALS, as developed within this thesis. Like in other domains, modelling must strike a balance between flexibility and interpretability. Classic challenges from longitudinal data include handling multiple outcomes simultaneously and dealing with unevenly spaced visits. However, in the context of chronic diseases, such as ALS, an additional challenge arises: the need to realign patients' trajectories, a task uniquely addressed by data-driven disease progression models. Survival modelling also presents its own set of challenges, primarily centred around censoring. Biologically, longitudinal and survival processes often exhibit interdependencies, prompting the development of joint models to estimate both concurrently. While various types of associations between these processes have been proposed, existing joint models fail to address the specificity of chronic diseases adequately. As a result, there is a pressing need for joint data-driven disease progression models tailored for this challenge.

Contents

1.1 Introduction

Many different regression models exist to analyse the occurrence of an event (survival data) and repeated measures at different visits (longitudinal data). From a biological perspective, these two processes are often associated and can inform each other. Thus, modelling both data together can reduce biases (Lu et al., 2023). Models that jointly optimise a longitudinal and a survival submodel have been developed for this purpose. These models are known as joint models.

In the context of clinical data, such models face the challenge of being as flexible as possible while keeping good interpretability. They are of specific interest in ALS, as an improvement in both survival and functional decline must be shown for treatment approval, according to FDA, 2019.

The following chapter details different existing longitudinal and survival models and how they were linked together to create joint models. Each part describes various applications of these models in the context of ALS.

The current state-of-the-art deliberately avoids the development of deep learning methods. Deep learning methods have demonstrated great performances in longitudinal data, especially in imaging (Valliani, Ranti, and Oermann, 2019) and natural language processing (Locke et al., 2021), but this thesis did not analyse any of them. Moreover, these algorithms may perform poorly when applied to clinical scores (Maheux et al., 2023; Christodoulou et al., 2019). They were also developed for survival analysis (Roblin, Cournède, and Michiels, 2024; Lee, Yoon, and Schaar, 2020; Lee et al., 2018; Jarrett, Yoon, and Schaar, 2019; Biganzoli et al., 2006) and joint modelling (Bull et al., 2020; Soleimani, Hensman, and Saria, 2017; Lee, Yoon, and Schaar, 2020). But overall, due to interpretation difficulties, overfitting issues and complex hyperparameters setting (Grollemund et al., 2019), they are often less practical (Kantidakis et al., 2023). Thus, despite their stringent assumptions, parametric regression models still have great potential in clinical applications (Westeneng et al., 2018).

1.2 Longitudinal models

1.2.1 Challenges from the data

Longitudinal data are repeated measures, for one patient, of one or several outcomes at different visits. Compared to cross-sectional data (one observation per patient), they enable to measure the long-term effect of covariables such as sex or characteristic of disease onset (Farrington, 1991). Their main challenge comes from the modelisation of the intra-patient correlation (several visits per patient), which needs to be taken into account to avoid misinterpretation of the data (Locascio and Atri, 2011). Then, different types of outcomes can be modelled. In this thesis, the outcomes studied were considered continuous. Literature dealing with the modelling of non-continuous outcomes is thus not further described (Ten Have et al., 1998; Liu and Hedeker, 2006; Saulnier et al., 2022; Poulet and Durrleman, 2023b).

For chronic diseases, as described in Introduction, longitudinal data often result in short-term heterogeneous irregular outcomes (Young et al., 2024). The trajectories must thus be realigned to reconstruct a comprehensive disease progression. Symptoms onset are often used to do so, but are not reliable in such diseases, as the progression of biomarkers often starts years before the manifestation of the first symptoms (Jack et al., 2010; Peter et al., 2017).

All the models described below, associated with the challenges they tackle, are summarised in Table 1.1.

1.2.2 Longitudinal models

Generalized Estimating Equation models (Liang and Zeger, 1986), have a "top-down" approach to estimate intra-patient correlation: they estimate directly the mean relations on the one hand, while separately dealing with covariance structure among the observations within subjects (Locascio and Atri, 2011). These models require equally spaced visits (Locascio and Atri, 2011) only give access to marginal models and not to individual trajectories, which makes them less used than other models (Gibbons, Hedeker, and DuToit, 2010).

To handle not regularly spaced visits, mixed-effect models were developed (Laird and Ware, 1982). Such models have a "bottom-up" approach and infer individual specificity to get the population one (Locascio and Atri, 2011). They are composed of two types of parameters, fixed effects, to describe the average population progression, and random effects, to capture the individual variability around the average disease progression. Linear Mixed-effects Models (LMM) are the simplest, but they are often inadequate because changes in clinical outcomes are non-linear over long periods of time.

The most common type of Non-Linear Mixed effects is the Generalised Linear Mixed-effect Model (GLMMs), a direct extension of LMM, which enables the estimation of non-linear progression (McCulloch, Searle, and Neuhaus, 2008). Even when non-linear, those models are fully parametric and might not be flexible enough to capture complex progressions.

Keeping the idea of individual parameters and irregular visits, ordinary differential equation models, more flexible than mixed-effect models, were developed. They are progression models that estimate the derivative of an outcome trajectory model, at the population level, from the partial trajectories of the observed patients (Lahouel et al., 2023). This derivative is then integrated to estimate the average progression and each initial condition of patients can be seen as individual parameters (as random effects in a mixed-effect model).

Finally, for even more flexibility, Random Forest models (Breiman, 2001) were designed. They are an ensemble model, that uses multiple decision tree models to predict an outcome, allowing to reduce the overfitting of decision trees used alone. Nevertheless, the flexibility of these models comes at the cost of interpretability, and if decision trees are easy to interpret, Random Forest models lose that possibility.

1.2.3 Data-driven disease progression models

Data-driven disease progression models are models designed "to learn long-term disease biomarker timelines for chronic diseases from short-term data without requiring prior knowledge of an individual's disease stage" (Young et al., 2024). Doing so, they are perfectly suited to tackle the challenges of data from chronic diseases. Here, we will only focus on phenomenological data-driven progression models that capture common trajectories of disease biomarkers as a continuous function of the data-driven time axis (Young et al., 2024). Note that pathophysiological data-driven progression models, focusing more on biological hypothesis testing and insight, also exist (Young et al., 2024), but were not considered in this thesis as no biological hypothesis was stated.

To improve the temporal resolution of existing GLMMs, data-driven mixed-effect models use a latent variable representative of the stage of the disease to realign trajectories. In that sense, disease progression score models were developed (Jedynak et al., 2012; Donohue et al., 2014) and extended (Bilgel et al., 2016; Bilgel and Jedynak, 2019; Raket, 2020). Other models created a latent disease age to realign patient observations (Li et al., 2017; Iddi et al., 2018).

The Spatiotemporal model is an extension of latent disease age models using geometry to operate in a high dimension (Durrleman et al., 2013; Schiratti et al., 2015; Schiratti, 2017). This model showed good performances in neurodegenerative disease by outperforming the 56 alternative methods for predicting cognitive decline in Alzheimer's disease in the TADPOLE challenge (Marinescu et al., 2019; Koval et al., 2021). This model was also extended to images (Lorenzi et al., 2019; Sauty and Durrleman, 2022; Abi Nader et al., 2020) and ordinal data (Poulet and Durrleman, 2023b).

To give more flexibility in modelling longitudinal outcomes, data-driven disease progression models using ordinary differential equation models were designed (Samtani et al., 2013; Villemagne et al., 2013; Jack et al., 2013; Oxtoby et al., 2014). Nevertheless, they do not estimate a disease timeline common to all outcomes and provide several disease stages for each individual (Young et al., 2024).

Table 1.1: Longitudinal models with their modelling challenges

Legend: +: the model addresses the challenge, +/-: the model partially addresses the challenge, -: the model does not address the challenge, Generalised Linear: Generalised Linear Mixed-effects Models, Latent disease age: Mixed effect models with a latent disease age, Spatiotemporal: Spatiotemporal model, Data-driven: Data-driven disease progression models here the Ordinary Differential Equation ones, Individual effects: individual patient specificity is captured by the model, High dimension: a large number of longitudinal outcomes can be modelled at the same time, Interpretability: the model is easy to interpret, Flexibility: the outcome progression function is very flexible, Not equally spaced visits: the model can handle not equally spaced visits easily without the need for imputations, Common timeline: the model can realign partial patients trajectories on a common timeline.

		Classic challenges					Chronic challenges
		Individual effects	High dimension	Interpretability	Flexibility	Not equally spaced visits	Common timeline
Mixed-effect	Generalised Linear	+	+/-	+	-	+	-
	Latent disease age	+	+/-	+	-	+	+
	Spatiotemporal	+	+	+	-	+	+
Ordinary Differential Equation	Classic	+	+/-	+	+/-	+	-
	Data-driven	+	+/-	+	+/-	+/-	+
Generalized Estimating Equation		-	+/-	+	-	-	-
Random Forest		-	+	-	+	-	-

Table 1.2: Survival models with their modelling challenges

Legend: +: the model addresses the challenge, +/-: the model partially addresses the challenge, -: the model does not address the challenge, Flexible hazard: the hazard function is flexible, Covariate effects: covariates can have several types of impact on survival, Interpretability: the model is easy to interpret, Individual effects: individual patient specificity is captured by the model.

	Flexible hazard	Covariates effect	Interpretability	Individual effects
Cox proportional hazard	+	-	+	-
Accelerated Failure Time	-	+/-	+	-
Survival Frailty	-	+/-	+	+
Random Survival Forest	+/-	+	-	-

1.2.4 ALS applications

ALS functional rating scale (ALSFRSr) (Rooney et al., 2017), is the main outcome used to monitor ALS progression and was invariably used in ALS modelling (Grollemund et al., 2019). Many linear mixed-effects models were used to analyse the heterogeneity of ALS by describing the impact of covariates (disease onset site, age, baseline blood markers) on the functional decline (ALSFRSr) (Visser et al., 2007; Atassi et al., 2014; Rooney et al., 2017) and the vital capacity (Atassi et al., 2014). Some GLMM were also used to analyse fMRI (Mejia et al., 2022). Random Forests enabled to model the vital capacity progression (Jahandideh et al., 2018). In line with these studies, we analysed, in chapter 2, the interaction of sex and onset site on the longitudinal progression using the Spatiotemporal model.

In 2015, a large data challenge on ALS, Prize4Life, aimed to find the best algorithm to predict prognosis to refine patient enrolment in clinical trials (Küffner et al., 2015). One task was to predict the slope of ALSFRS changes between 3-12 months with ALSFRS measured between 0-3 months on the large PRO-ACT database. Many of the participants were using Random Forest (Hothorn and Jung, 2014; Ko et al., 2014) and some argue that they better generalise than GLMMs (Taylor et al., 2016). The results of this challenge could be seen as state-of-the-art in the prediction of ALS prognosis and will be further referred to in chapter 3.

1.3 Survival models

1.3.1 Challenges from the data

When studying the occurrence of an event, different estimands are of interest:

- Survival ($S(t)$): the probability that the event occurred later than a specified time t: $S(t) = P(T > t)$,
- Hazard ($h(t)$): the probability that the event occurred at a given time t, conditional on survival until time t or later: $h(t) = \lim_{dt \to 0} \frac{P(t \leq T < t+dt)}{dtS(t)}$
- Cumulative Incidence Function ($CIF(t)$): the probability that an event of a given type occurs at or before time t: $CIF(t) = P(T \leq t)$.

The main challenge in computing these quantities is that the event is not always observed. Not taking into account this missingness, named censoring in survival analysis, would lead to biased estimations (Bouaziz, n.d.). Censoring can occur at different times compared to the event: 1) left censoring, when the event occurs before the window of follow-up; 2) right censoring, when the event occurs after the window of follow-up and 3) interval censoring, when the event occurs between two visits (Leung, Elashoff, and Afifi, 1997). The association between the censoring process and the event is then the major difficulty. Most of the models assume a non-informative censure (also known as the independent censoring assumption), meaning that the hazard of the event for a patient censored at T_C is the same as the one for the remaining patients after T_C (Fleming and Harrington,

2013; Jackson et al., 2014). Jackson et al., 2014 remind that this assumption is slightly weaker than the assumption of statistical independence between event and censoring conditionally on covariates (missing at random data (Rubin, 1976; Nahhas, 2024)) that is often used. When modelling events with informative censoring, the main challenge is to model the source of censoring. This source can be identified, although partially observed, and the task results in modelling several competing events. In this thesis, we considered only right censored data with non-informative censoring or with competing risks.

Note that recent pseudo-observation techniques have provided a solution to the censorship issue in survival data, allowing them to be treated similarly to longitudinal data without censorship concerns (Andersen and Perme, 2010). However, this technique requires a restriction to the study of specific time points. This is not always appropriate and will not be discussed further here.

1.3.2 Modelling an event with non-informative right censoring

When modelling an event with non-informative right censoring, there is a one-to-one mapping between the modelling of $h(t)$ and $CIF(t) = 1 - S(t)$ (Andersen et al., 2012), in such case, survival is often the main output of interest. Different methods exist to model the occurrence of events and are summarised jointly with the challenges they tackle in Table 1.2.

The most known survival model is the Kaplan-Meier estimator (Kaplan and Meier, 1958). It is a non-parametric model, thus very flexible, that describes a cohort, but is limited in its ability to estimate survival adjusted for covariates.

Hazard-based regression models were partly designed to improve the description of covariates' impact on the hazard (Rubio et al., 2019). The most famous is the Cox Proportional Hazard model (Cox, 1972), with a proportional effect of covariables on the hazard through the multiplication of a baseline hazard. Its main interests are that it does not require estimating the baseline hazard function when used alone, which enables some flexibility in the hazard. Nevertheless, for some applications (visualisation, joint modelling), a flexible parametric version of the Cox model was also developed by Royston and Parmar, 2002. Even if the proportional impact of covariates on the hazard is easy to interpret, this assumption is often violated.

Accelerated Failure Time (AFT) models enable to describe a non-proportional impact of covariates on the hazard (Kalbfleisch and Prentice, 2002). In these models, covariates directly affect survival time, but in doing so, modelling the hazard function is mandatory, making them less used for survival analysis alone. Besides, the results obtained from these models may be harder to interpret.

Frailty models alleviate the assumption of the Cox model by capturing an additional proportional effect of unobserved covariates using a random effect (Hougaard, 1995). However, to calculate the random effect, correlated survival data (observations clustered into groups or recurrent events) are needed, which are not always available.

Finally, to access even more flexibility in modelling, Random Survival Forest models were adapted using a measure of entropy suited for survival (log-rank test) (Ishwaran et al., 2011; Jiang, Xie, and Colditz, 2020; Lin, Li, and Luo, 2021). Nevertheless, as for the longitudinal models (see section 1.2.2), this is at the cost of interpretability.

1.3.3 Modelling an event with competing risks

When studying competing events, there is no longer a one-to-one mapping between the hazard and the CIF for an event l as the cumulative incidence takes into account the survival from all causes (CIF_l is distinct from $1-S_l(t)$) (Zhang, Zhang, and Scheike, 2008; Andersen et al., 2012. In such a context, the CIF is often the main output of interest.

From there, historically, two main approaches exist to model CIF (Zhang, Zhang, and Scheike, 2008). The first one consists of modelling the hazard with different cause-specific hazard functions. The most famous approach is the one using the proportional hazard Cox approach (Prentice and Gloeckler, 1978; Cheng, Fine, and Wei, 1998) but others using the Aalen additive hazard model (Aalen, 1989) were also developed. With the cause-specific model, each event is modelled separately. In the case of the Cox version, the hazard ratio can thus still be extracted (Andersen et al., 2012). For these reasons, it is particularly relevant for the study of disease aetiology. Frailty models were extended using cause-specific modelling for competing risks but still require correlated survival data (Gorfine and Hsu, 2011). Nevertheless, with the cause-specific models, the impact of covariates directly on the CIF is hard to interpret.

The second one consists in modelling the distribution hazard function instead of the hazard itself, which enables a more direct effect of covariates on the cumulative incidence function. The proportional Cox version of such an approach is often known under the Fine and Gray model (Fine and Gray, 1999). This method is useful to reflect the effect of covariates on the ordering of the cumulative incidence curves but only gives access to sub-distribution HRs which have no easy interpretation (Zhang, Zhang, and Scheike, 2008; Andersen et al., 2012). In addition, as it models jointly multiple causes of events this method is not recommended for causal inference (Allison, 2018).

A more recent approach, Vertical Modelling, separates the problem into two different quantities: the total hazard (to access the probability of being free from all causes) and the relative cause-specific hazards (Nicolaie, Van Houwelingen, and Putter, 2010). Although in terms of probability decomposition, it may seem more intuitive, two distinct functions must be specified and the same problem as with the Fine Gray model for causal inference might occur (Allison, 2018).

Random Survival Forests were also extended with a suited entropy measure (Fine and Gray test) and give more flexibility in modelling still at the cost of interpretability (Ishwaran et al., 2014; Devaux et al., 2023).

1.3.4 ALS applications

In ALS, death has been the first outcome studied, before functional decline. Some studies tried to characterise the heterogeneity of the disease progression using a Cox proportional hazard model to understand the impact of covariates on survival (Murdock et al., 2021).

The same kind of model was also used to demonstrate the impact of treatments: for riluzole (Ad, Rpa, and Lh, 2019; Zoccolella et al., 2007; Traynor et al., 2003; Georgoulopoulou et al., 2013), gastrostomy (Hesters et al., 2020; Vergonjeanne et al., 2020) or NIV (Kleopa et al., 1999; Lo Coco et al., 2006; Bourke et al., 2006).

The Prize4Life challenge (Küffner et al., 2015) also included the prediction of the probability of death within 0–12 months, 0-18 and 0–24 months from trial onset. Several Random Forests were used (Beaulieu-Jones, Greene, and Pooled Resource Open-Access ALS Clinical Trials Consortium, 2016; Pfohl et al., 2018) including the winner (Huang et al., 2017). To give access to such predictions to neurologists, Westeneng et al., 2018 have developed an online tool using the Cox Royston-Parmar model (Royston and Parmar, 2002): `http://www.encalssurvivalmodel.org/`.

There have been relatively few studies investigating competing risks in ALS. Even though vital capacity was early identified as a predictor of disease progression (Schiffman and Belsh, 1993) and proved relevant in the Prize4Life challenge (Küffner et al., 2015), few studies model None-Invasive-Ventilation introduction (Thakore et al., 2019).

1.4 Joint models

1.4.1 Challenges from the data

From a biological perspective, longitudinal and survival processes are often associated, and modelling both data together reduces biases (Lu et al., 2023). Joint models were thus developed by connecting a survival model with a longitudinal submodel. The principal challenge from the data is to well model the association of the two processes and many different structures have been proposed offering different levels of flexibility to capture individual variability in the survival submodel.

Different models have been published. They are described below, associated with the challenges they tackle, and summarised in Table 1.3.

Note that some methods, such as landmarking or two-stage modelling (Bull et al., 2020), use the output of longitudinal models to inform survival models, in a two-step manner. Nevertheless, joint modelling has the advantage of carrying forward the uncertainty and reducing biases in the hazard ratio (Sweeting et al., 2017). Thus, these methods will not be further described here.

1.4.2 Latent class joint models

The Latent class joint model, originally developed by (Lin et al., 2002; Proust-Lima et al., 2009; Proust-Lima et al., 2014), groups similar patients together in a latent class with a specific value of parameters per group. The survival function has thus different fixed effects for each latent class. By aggregating patients, this approach helps to better understand the heterogeneity, even though the meaning of the different classes remains unknown. It allows calculating the probability of an individual belonging to a particular class but may result in some patients being almost equitably distributed.

As a longitudinal submodel, they are all built on GLMM for each latent class (mixture model). The survival submodel was originally developed for events with non-informative right censoring (Proust-Lima et al., 2009) using a Cox proportional association for the covariates. Now, many extensions exist, enabling competing risks with cause-specific models (Blanche et al., 2015; Proust-Lima et al., 2014; Proust-Lima, Dartigues, and Jacqmin-Gadda, 2016). These developments are associated with a software package developed in R (Proust-Lima, Philipps, and Liquet, 2017).

1.4.3 Shared random effects joint models

Shared random effect models associate with each individual a random effect and could be seen as a direct extension of the longitudinal mixed effect models (Rizopoulos and Ghosh, 2011). The survival function thus integrates a proportional effect of the functions of the random effects of the longitudinal model, allowing more flexibility in capturing its variability. They are the most used type of joint models (Sudell, Kolamunnage-Dona, and Tudur-Smith, 2016) as they enable capturing individual variability in the survival function. Their main limitation is that by including predictors of the longitudinal outcome in the survival model, they mostly focus on how longitudinal outcomes impact survival.

As the other joint models, they are all built on GLMM. For the survival submodel, the original model used a Cox proportional association for the longitudinal outcome (Rizopoulos and Ghosh, 2011) and then others were developed using AFT (Tseng, Hsieh, and Wang, 2005).

Extensions that model competing events mainly use a cause-specific formulation with Cox proportional hazard (Elashoff, Li, and Li, 2007; Andrinopoulou et al., 2017) or AFT (Hao et al., 2024). More recently, a joint model using the Fine Gray approach was developed (Thomadakis et al., 2024). A recent extension combined latent class to random effect in a high dimensional context (16 longitudinal outcomes) (Nguyen et al., 2023). In parallel, different estimation software packages were implemented for joint models based on GLMMs (Rizopoulos, Papageorgiou, and Miranda Afonso, 2024; Rustand et al., 2024).

1.4.4 Frailty joint models

Frailty joint models were developed (Rondeau, Mazroui, and Gonzalez, 2012; Masci, Spreafico, and Ieva, 2023), extending the frailty survival models described in section 1.3.2. They enable capturing an additional random effect through the correlated survival data, compared

to the shared random effects model. Nevertheless, as for frailty survival models, such correlated survival data are not always available. As with other joint models, they are grounded on GLMM and have been extended for competing events (Huang, Li, and Elashoff, 2010; Etzkorn et al., 2024). A dedicated software package was also created (Rondeau, Mazroui, and Gonzalez, 2012).

TABLE 1.3: Joint models with their modelling challenges

Legend: +: the model addresses the challenge, +/-: the model partially addresses the challenge, -: the model does not address the challenge, Individual effects: individual patient specificity is captured by the model, High dimension: a large number of longitudinal outcomes can be modelled at the same time, Common timeline: the model can realign partial patients trajectories on a common timeline, Link between submodels: describe how the longitudinal and the survival submodel are linked inside of the joint model.

	Classic challenges			Chronic challenges
	Individual effects	High dimension	Link between submodels	Common timeline
Latent class	+/-	+/-	Parameters per group of patients	-
Shared random effects	+	+/-	Impact of longitudinal outcomes on survival	-
Frailty	+	+/-	Extract information from correlated survival information	-

1.4.5 ALS Application

Most of the studies used a two-step analysis to analyse longitudinal data jointly with survival data (Bull et al., 2020; Murdock et al., 2021; Haverkamp, Appel, and Appel, 1995; Jablecki, Berry, and Leach, 1989; Norris et al., 1993; Atassi et al., 2014; Elamin et al., 2015; Reniers et al., 2017). Only a few studies have used joint models in different contexts: Sun et al., 2020 studied the impact of a longitudinal outcome (blood markers) on an event (death), Rooney et al., 2019 studied the impact of a covariate (the gene C9orf72) on both a longitudinal outcome (ALSFRSr) and an event (death) and Kyheng et al., 2021 stratified the population thanks to both a longitudinal outcome (ALSFRSr) and an event (death). Joint models have also shown a potential to demonstrate treatment effects (Eijk et al., 2019) and were applied to the analysis of the effects of ceftriaxone (Wang and Luo, 2019). Nevertheless, joint models with competing risks in ALS have not been as much used as in other diseases.

1.5 Discussion

To model longitudinal data, mixed effect models and the Ordinary Differential Equation models are interesting because they give access to both population and individual levels with good interpretability. From there, data-driven progression mixed effect models seem better suited to our neurodegenerative context as they enable us to realign patients' trajectories on a common timeline. More specifically, the Spatiotemporal model that was originally designed for high dimensional longitudinal data (Durrleman et al., 2013) is particularly interesting because it has already been widely used in the context of neurodegenerative diseases.

An important number of the survival models are non-parametric or semi-parametric ones. However, when associated with a longitudinal submodel, the survival submodel must be parametric (Rizopoulos, 2012), which reduces the interest of the Cox model compared to the AFT model for events with non-informative right censoring. For events with competing risks the cause-specific model is recommended to describe disease progression heterogeneity. Whatever the type of censure, frailty models are the only survival models to integrate individual parameters to model unobserved covariates but require correlation in the survival data.

Joint modelling enables to bring more flexibility to capture survival variability. The shared random effect models are the most used as they enable to capture individual variability in survival, thanks to random effects, without the need for correlated survival data. Nevertheless, whatever the extension, all these joint models rely on GLMMs, necessitating the use of reference times, which is not reliable in our context. In that sense, a data-driven disease progression joint model, that might be able to handle a high dimension of features is missing and may bring a different way to analyse the data (Hickey et al., 2016; Alsefri et al., 2020). The objective of this thesis was thus to create a Joint data-driven model to handle events with non-informative right censoring or competing risks jointly with multiple longitudinal

outcomes. Note that in the meantime, other efforts on joint modelling with temporal recalibration (Saulnier, Proust-Lima, and Samier, 2023) and joint models grounded on non-linear mixed effects models more complex than GLMM (Lavalley-Morelle et al., 2024) were also conducted.

Chapter 2

Quantifying ALS heterogeneity with a longitudinal Spatiotemporal model

In this chapter, we present the existing longitudinal Spatiotemporal model and use it to study the impact of sex and onset site (spinal or bulbar) interaction on the progression of ALS. We selected patients from the PRO-ACT database and divided them into four subgroups depending on their sex and onset site. We used one Spatiotemporal model to investigate the combined changes of the four sub-scores of the revised ALS Functional Rating Scale (ALSFRSr), the forced vital capacity (FVC), and the body mass index (BMI). We then compared the progression speed, the estimated reference time, and the relative progression of the outcomes across each subgroup. We included 1,438 patients: 51% spinal men, 12% bulbar men, 26% spinal women, and 11% bulbar women. We showed a significant influence of both sex and onset site on the ALSFRSr progression. The BMI decreased 8.9 months earlier (95% CI = [3.9, 13.8]) in women than men, after correction for the onset site. Among patients with bulbar onset, FVC was impaired 2.6 months earlier (95% CI = [0.6, 4.6]) in women. Using the Spatiotemporal model, we showed that sex and onset site are important drivers of the progression of motor function, BMI, and FVC decline.

This chapter was published in the Journal of Neurology. See (Ortholand et al., 2023).

Contents

2.1 Introduction

ALS shows an important phenotypic variability, particularly regarding the site of disease onset and progression (Swinnen and Robberecht, 2014). It is well documented that the forms starting with dysarthria and dysphagia (bulbar onset) or with limb weakness (spinal onset) differ in terms of incidence (Swinnen and Robberecht, 2014), progression speed (Tysnes, Vollset, and Aarli, 1991; Brown and Al-Chalabi, 2017; Brooks, 1996) and survival duration (Tysnes, Vollset, and Aarli, 1991).

Another major component of disease heterogeneity is the marked sexual dimorphism. Several epidemiological studies showed a lower incidence of ALS in women compared to men (Tysnes, Vollset, and Aarli, 1991; Logroscino et al., 2010) as well as differences in the onset site and age at onset (Chiò et al., 2020).

However, the interaction of sex and onset site on disease trajectory is not fully described. First, most results in the literature come from cross-sectional studies (Logroscino et al., 2010; Chiò et al., 2020), although longitudinal studies are needed to measure the long-term effect of clinical or demographic variables, such as age or sex (Farrington, 1991). Second, studies mainly concerned survival (Tysnes, Vollset, and Aarli, 1991) and did not take into account how the interaction of age and disease impacts the progression and the presentation of the disease.

In chapter 1, we have seen that mixed-effects models could be of great help when analysing longitudinal data. More precisely, we have decided to take advantage of the existing Spatiotemporal model (Schiratti et al., 2015; Schiratti, 2017), well suited to realign the partial follow-up of patients in the context of chronic diseases.

After presenting the Spatiotemporal model in more detail, we used it on a large longitudinal dataset to investigate the interaction of sex and onset site on the progression of the following clinical proxies: BMI (Moglia et al., 2019), Forced Vital Capacity (FVC) (Daghlas, Govindarajan, and Pooled Resource Open-Access ALS Clinical Trials Consortium, 2021), and the revised ALS functional rating scales (ALSFRSr) (Rooney et al., 2017).

2.2 Model specifications

2.2.1 Notations

In the following, we consider N patients, associated with longitudinal data: repeated measures of K given outcomes noted y_k. Each patient i is followed for n_i visits. For each visit j, we denote $t_{i,j,k}$ the age at the visit, and $y_{i,j,k}$ the value of the outcome k for the patient i at this visit j.

2.2.2 Intuition and Riemannian geometry background

The Spatiotemporal model was first introduced by Schiratti et al., 2015 and was more broadly used for longitudinal modelling of neurodegenerative diseases (Schiratti, 2017; Koval, 2020) as described in chapter 1. As presented in Figure 2.1 the model draws a parallel

between a clinical and a Riemannian point of view of the disease progression. The idea is to see the variability of the disease progression as a Riemannian manifold where the longitudinal observations $y_{i,j,k}$ are aligned in an individual trajectory γ_i that traverses the manifold.

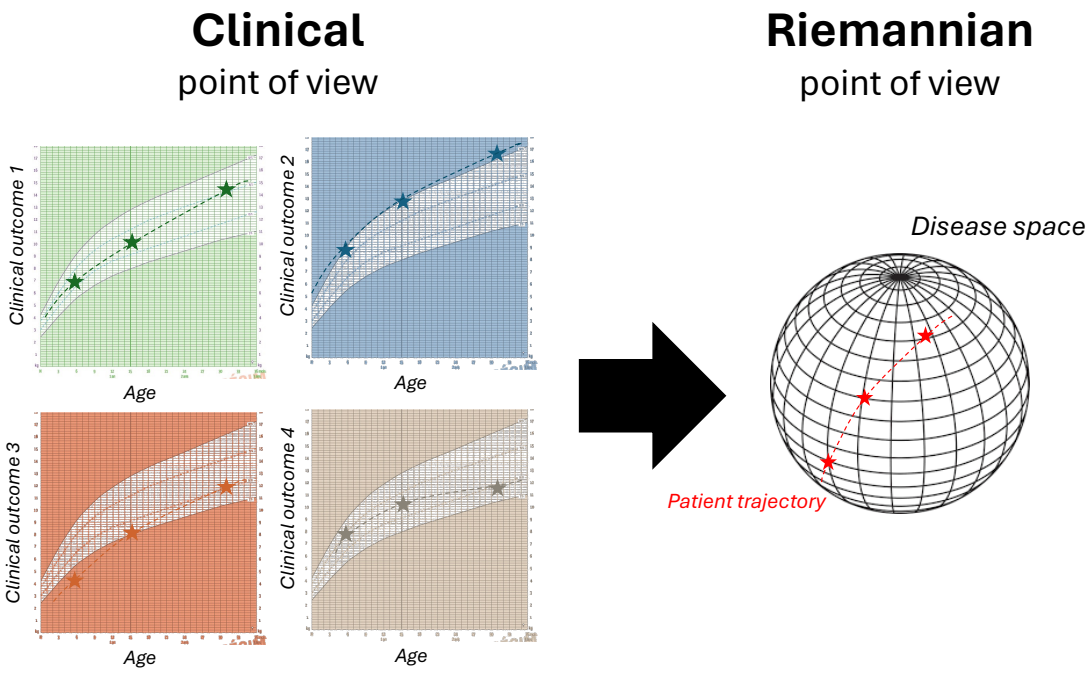


FIGURE 2.1: From clinical to Riemannian point of view
On the left, the progression of four clinical outcomes for one patient is represented depending on the age of the patient. The graph displays the individual progression of one patient on a grid detailing the typical progression of the disease, as it is done in health diaries for BMI curves. This represents how a clinician is used to see the progression of the patient. On the right, the same patient progression is represented but this time in a disease space (manifold) built thanks to the knowledge extracted from the four clinical outcomes. This represents the Riemannian point of view of the progression of the patient.

The shape of the disease progression (linear, logistic ...) is defined by the choice of the Riemannian metric applied to the manifold. For instance, the manifold $\mathbb{R}^n$ equipped with Euclidean metric gives straight lines trajectories and thus straight lines disease progression. As we studied mainly clinical scores, with curvilinearity, and potential floor or ceiling effects (Gordon et al., 2010), we selected a metric that enables modelling the logistic progression of the outcomes.

To separate an average disease progression from the individual progression, a mixed-effects model structure is added to the trajectories. Any trajectory γ (geodesic) can be defined by the two parameters of its initial condition at a time t_0: the initial position $\gamma(t_0) = p$ and the initial speed $\dot{\gamma}(t_0) = v_0$. The average trajectory γ_0 is thus parametrized by its initial conditions (t_0, p, v_0). This average longitudinal process is further described in section 2.2.3.3.

From there, the individual trajectory $\gamma_i(t)$ could be defined. First, a temporal variation is enable with varying degrees of individual earliness τ_i and speed e^{ξ_i}, using a latent disease age $\psi_i(t)$. In terms of initial definition, these variations could be seen as $(t_0 + \tau_i, p, v_0 e^{\xi_i})$. Subsequently, variation in terms of disease presentation, which corresponds to modifying the order of degradation of the various outcomes, is allowed. From a Riemannian point of view, the trajectory is spatially adjusted playing on the initial position p. It is done thanks to a vector in the tangent space of the trajectory that modified the trajectory in the sense of the Exp-parallelisation. All these individual parameters are further described in section 2.2.3.2.

2.2.3 The longitudinal Spatiotemporal model

2.2.3.1 Mixed effect models formalism

The longitudinal process modelled, $\gamma_{i,k}(t_{i,j,k})$, is the progression of the k^{th} outcome of the patient i, measured by $y_{i,j,k}$ at each time $t_{i,j,k}$ for each visit j of the patient i. t is also indexed by k so that the model can handle missing values on some outcomes. The longitudinal process is estimated with a Gaussian noise for each outcome compared to the measure: $\epsilon_k \sim \mathcal{N}(0, \sigma_k)$ with σ_k the standard deviation of the noise of the outcome k, which gives: $y_{i,j,k} = \gamma_{i,k}(t_{i,j,k}) + \epsilon_{i,j,k}$. Note that we assume that all the noises of the visits are independent.

To describe the model, we will further use the formalism of mixed-effects models: with fixed effects (population level parameters) and random effects (individual level parameters indexed by i) (see section 1.2.2).

2.2.3.2 Random effects: temporal and spatial individual variability

The strength of the Spatiotemporal model is to disentangle temporal from spatial variability.

Temporal variability First, temporal variability is allowed with variations on individual progression earliness and speed. It is done by mapping the chronological age of a patient t into a latent disease age $\psi_i(t)$, representative of the disease stage of the patient. Using the formalism described before, it can be written as :

$$\psi_i(t) = \quad e^{\xi_i}(t - \tau_i) + t_0 \tag{2.1}$$

where e^{ξ_i} is the speed factor of patient i, τ_i is its individual estimated reference time and t_0 is the population estimated reference time. $(\tau_i - t_0)$ can thus be seen as an individual time-shift compared to the population. Although the reference time is not the time of disease onset, it plays quite the same role: it is a state of the disease on which all the patients are realigned. The effects of the two temporal random effects on the progression are described in Figure 2.2 from both a clinical and a Riemannian point of view. The main advantage of this formalism is that the individual progressions are realigned on a given value of the outcomes and not on a reference onset time, as with Generalised Linear Mixed effects models (Schiratti et al., 2015), which might be more robust in our context.

Spatial variability To capture the disease presentation variability, spatial random effects, named the space-shifts $w_{i,k}$, are defined for each outcome to modify their order of degradation during the disease progression. Nevertheless, for identifiability reasons, the dimension of the space-shift space is reduced with an independent component analysis (ICA) decomposition using $N_s \leq K-1$ independent sources $(s_i)_{1\leq i\leq N_s}$, resulting in $w_i = As_i$, where A is the mixing matrix of the ICA decomposition. However, this definition does not guarantee the orthogonality of the space shift to the speed of progression v_0 (as in the Exp-parallelisation at $\frac{1}{1+g}$ from Riemannian geometry) which gives the identifiability. Thus, the matrix A is defined as a linear combination of vectors of an orthonormal basis, $(B_o)_o$, of the hyperplane orthogonal to $Span(v_0)$ (dimension $K \times (K-1)$): each column m of A is thus $A_m = \sum_{o=1}^{K-1} \beta_{o,m} B_o$ with β the matrix of coefficient (dimension $(K-1) \times N_s$) so that $A = (B\beta)^T$. (Schiratti et al., 2017). The effects of these spatial random effects on the progression are described in Figure 2.2.

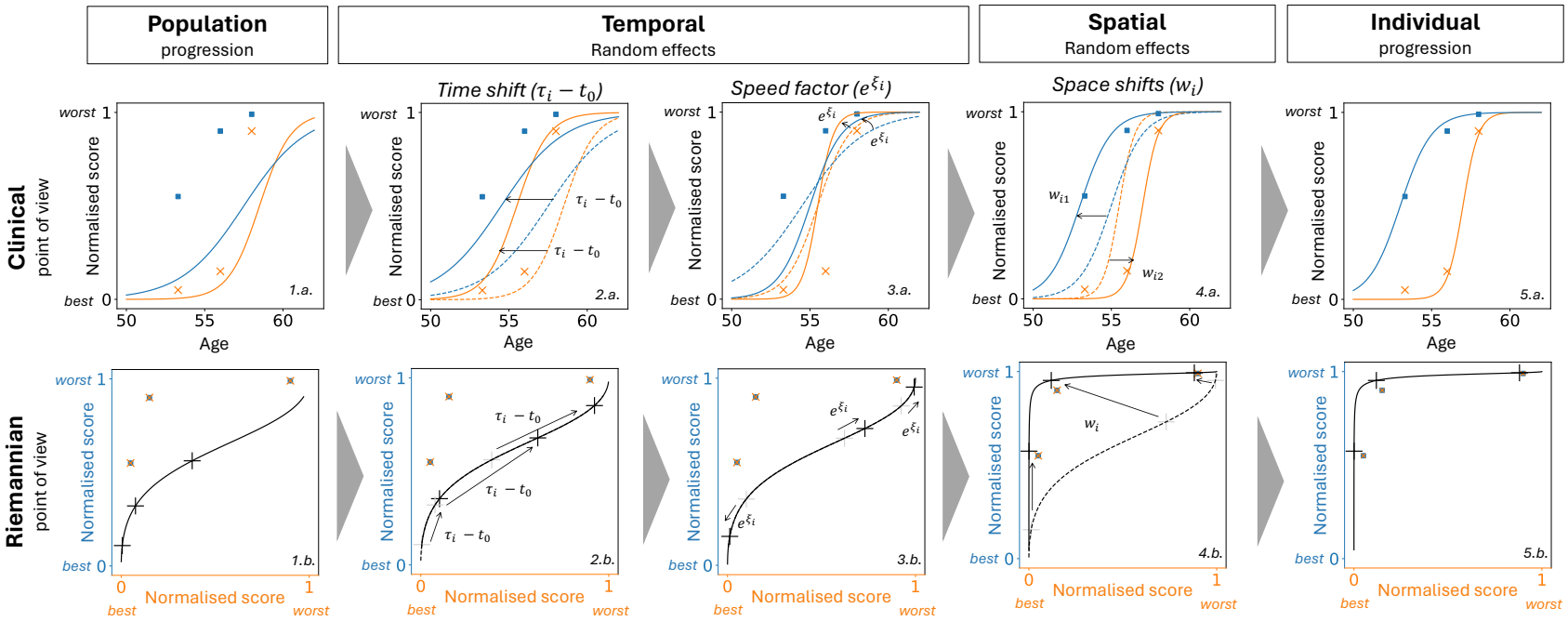


FIGURE 2.2: Temporal and spatial random effects: from population to individual progression (adapted from Koval et al., 2021)

This figure presents from two points of view (clinical and Riemannian) how the three types of random effects (two temporal and one spatial) enable to modify the population average progression to calibrate the patient observations. ***Clinical:*** *Two normalised clinical scores (blue and orange) (0: the healthiest value, +1: the maximum pathological change) depending on the age of the patients. The scatter represents the real observed values for one patient at different visits.* ***Riemannian:*** *The same two normalised scores are represented but this time depending on each other. The scatter represents the same real observed values as in the clinical version. The black cross on the curve corresponds to what is modelled at the visit ages of the patient.* ***Population progression (1.a., 1.b.):*** *Population average trajectory compared to the observed values of the patient.* ***Time Shift (2.a., 2.b.):*** *The progression starts earlier due to the individual estimated reference time, Clinical graph: the curves are shifted on the left, Riemannian graph: black crosses are shifted on the right following the trajectory (for the same age the patient is more advanced).* ***Speed factor (3.a., 3.b.):*** *The progression speed increases, Clinical graph: the curves become steeper, Riemannian graph: black crosses get further from each other on the trajectory (for the same time of follow-up a wider portion of the trajectory is observed).* ***Space Shift (4.a., 4.b.):*** *the blue curves progress before the orange curve, Clinical graph: the curves are shifted in opposite directions, Riemannian graph: most of the blue (resp. orange) score value is observed for an orange (resp. blue) value of 0 (resp. 1)* ***Individual progression (5.a., 5.b.):*** *The modelled curves fit the observations, Riemannian graph: the black crosses are close to the observed values.*

2.2.3.3 Fixed effects: trajectory shape

The modelling of the longitudinal process consists of computing the trajectory from the latent disease age defined in section 2.2.3.2. To model logistic curves, we used a metric $G(p) = \frac{1}{p^2(1-p)^2}$ in a manifold $(0, 1)$. For a K-dimensional dataset, we used the product manifold of this 1-dimensional manifold. The method to find the trajectory from the metric is further described in Koval, 2020 (p.169). From there, we got the average curve for an outcome k at time t:

$$\gamma_{0,k}(t) = \left(1 + g_k \times \exp(-\frac{(1+g_k)^2}{g_k}(v_{0,k}(t - t_0))\right)^{-1} \tag{2.2}$$

where t_0 is the population estimated reference time defined in section 2.2.3.2, $v_{0,k}$ is the speed of the logistic curves at t_0 and $p_k = \frac{1}{1+g_k}$ is the value of the modelled outcomes at t_0. We also got the individual curve for an outcome k (continuous between 0 and 1), an individual i at time t:

$$\gamma_{i,k}(t) = \left(1 + g_k \times \exp(-\frac{(1+g_k)^2}{g_k}(v_{0,k}(\psi_i(t) - t_0) + w_{i,k})\right)^{-1} \tag{2.3}$$

2.2.3.4 Wrap-up

To sum up, the model described can be resumed by the following formulas for a patient i and an outcome k:

$$\begin{cases} \psi_i(t) = e^{\xi_i}(t - \tau_i) + t_0 \\ w_i = As_i \\ \gamma_{i,k}(t) = \left(1 + g_k \times \exp(-\frac{(1+g_k)^2}{g_k}(v_{0,k}(\psi_i(t) - t_0) + w_{i,k})\right)^{-1} \end{cases} \tag{2.4}$$

Note that the model described here is not structurally identifiable (Lavielle and Aarons, 2016) but its identifiability comes with the statistical model.

2.2.4 Likelihood of the longitudinal Spatiotemporal model

2.2.4.1 Parameters structure

In the statistical model, all parameters of the structural model described in section 2.2.3.4 are latent parameters following a given distribution. We make a distinction between fixed and random effects. Random effects follow a Gaussian distribution for which all parameters are estimated using likelihood maximisation. To ensure the positivity of fixed effects, their log are estimated and written with a tilde (for instance, $\tilde{v}_0 = \log(v_0)$). Log fixed effects follow a normal distribution with the average parameters estimated using a log-likelihood maximisation and the standard deviation parameters defined as hyperparameters. To ensure identifiability, we set the mean parameter $\overline{\xi} = 0$, $t_0 = \overline{\tau}$, $\overline{s} = 0$, and the standard deviation parameters $\sigma_s = 1$. We end up with the following structure with the patients indexed by i and outcomes by k, the sources by m, and the dimensions of the hyperplane orthogonal to $Span(v_0)$ by o:

- Latent parameters (z):

- Latent fixed effects (z_{fe}): log fixed effects sampled from normal distributions parametrized by model parameters and hyperparameters $z_{fe} = \{\tilde{g}_k, \tilde{v}_{0,k}, \beta_{o,m}\}$,

$$\tilde{g}_k = \log(g_k) \sim \mathcal{N}\left(\overline{\tilde{g}_k}, \sigma^2_{\tilde{g}}\right) \quad \tilde{v}_{0,k} = \log(v_{0,k}) \sim \mathcal{N}\left(\overline{\tilde{v}_{0,k}}, \sigma^2_{\tilde{v}_0}\right) \quad \beta_{o,m} \sim \mathcal{N}\left(\overline{\beta}_{o,m}, \sigma^2_{\beta}\right)$$

- Latent random effects (z_{re}): sampled from normal distributions parametrized by model parameters $z_{re} = \{\xi_i, \tau_i, s_i\}$,

$$\xi_i \sim \mathcal{N}\left(\overline{\xi}, \sigma^2_{\xi}\right) \qquad \tau_i \sim \mathcal{N}\left(\overline{\tau}, \sigma^2_{\tau}\right) \qquad s_{i,m} \sim \mathcal{N}\left(\overline{s}, \sigma_s\right)$$

- Model parameters (θ): fixed effects estimated from log-likelihood maximisation $\theta = \{\sigma_\xi, \sigma_\tau, t_0, \overline{\tilde{g}_k}, \overline{\tilde{v}_{0,k}}, \overline{\beta_{o,m}}, \sigma_k\}$
- Hyperparameters (Π): set by the user $\Pi = \{\sigma_{\tilde{g}}, \sigma_{\tilde{v}_0}, \sigma_\beta\}$

2.2.4.2 Likelihood

The likelihood estimated by the model is the following:

$$p(y \mid \theta, \Pi) = \int_z p(y, z \mid \theta, \Pi) dz$$

$p(y, z \mid \theta, \Pi)$ can be divided into two different terms: data attachment, which represents how well the model describes the data y, and a prior attachment, which prevents over-fitting:

$$\log p(y, z \mid \theta, \Pi) = \quad \log p(y \mid z, \theta, \Pi) + \log p(z \mid \theta, \Pi)$$

The prior attachment term can be separated into two terms: two terms for the prior attachment of latent parameters (fixed and random). We end up with the following expression :

$$\log p(y, z \mid \theta, \Pi) = \quad \log p(y \mid z, \theta, \Pi) + \log p(z_{re} \mid z_{fe}, \theta, \Pi) + \log p(z_{fe} \mid \theta, \Pi)$$

The four different parts of the log-likelihood are described below associated with their different assumptions. Note that the total formula of the log-likelihood is available in appendix A.1.

Data attachment To model the longitudinal process, we assumed that patients and their visits are independent conditionally on random effects and that the noise of the process follows a Gaussian distribution. We thus got (from Koval, 2020 p.175):

$$\begin{aligned}\log p(y \mid z, \theta, \Pi) &= \sum_{i,j,k} \log p(y_{i,j,k} \mid z, \theta, \Pi) \\ &= \sum_{i,j,k} -\log\left(\sigma_k \sqrt{2\pi}\right) - \frac{1}{2\sigma_k^2} \left(y_{i,j,k} - \gamma_{i,k}\left(t_{i,j,k}\right)\right)^2\end{aligned}$$

Latent random effects priors attachment As patients were supposed independent of each other, we supposed that random effects were independent conditionally to z_{fe}, θ, and

Π. The regularization term associated, with $\overline{\xi} = 0$, $t_0 = \overline{\tau}$, $\overline{s} = 0$ and $\sigma_s = 1$, is then (from Koval, 2020 p.175):

$$\begin{aligned}
\log p(z_{re} \mid z_{fe}, \theta, \Pi) = & \sum_i \left(\log p(\tau_i \mid z_{fe}, \theta, \Pi) + \log p(\xi_i \mid z_{fe}, \theta, \Pi) + \sum_m^{N_s} \log p(s_{i,m} \mid z_{fe}, \theta, \Pi) \right) \\
= & - N \log\left(\sigma_\tau \sqrt{2\pi}\right) - \frac{1}{2\sigma_\tau^2} \sum_i (\tau_i - t_0)^2 \\
& - N \log\left(\sigma_\xi \sqrt{2\pi}\right) - \frac{1}{2\sigma_\xi^2} \sum_i (\xi_i - \overline{\xi})^2 \\
& - N N_s \log\left(\sigma_s \sqrt{2\pi}\right) - \frac{1}{2\sigma_s^2} \sum_i \sum_m^{N_s} (s_{i,m})^2
\end{aligned}$$

Latent fixed effects prior attachment Each latent fixed effect is independently sampled from a posterior distribution. The regularization term associated is then (from Koval, 2020 p.175):

$$\begin{aligned}
\log p(z_{fe} \mid \theta, \Pi) = & \sum_k \left(\log p(\tilde{g}_k \mid \theta, \Pi) + \log p(\tilde{v}_{0,k} \mid \theta, \Pi) \right) \\
& + \sum_{o,m} \log p(\beta_{o,m} \mid \theta, \Pi) \\
= & - \sum_k \log\left(\sigma_{\tilde{g}} \sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{g}}^2} \left(\tilde{g}_k - \overline{\tilde{g}}_k\right)^2 \\
& - \sum_k \log\left(\sigma_{\tilde{v}_0} \sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{v}_0}^2} \left(\tilde{v}_{0,k} - \overline{\tilde{v}}_{0,k}\right)^2 \\
& - (K-1) N_s \log(\sigma_\beta \sqrt{2\pi}) - \frac{1}{2\sigma_\beta^2} \sum_{o,m} (\beta_{o,m} - \overline{\beta}_{o,m})
\end{aligned}$$

2.2.5 Estimation of the longitudinal Spatiotemporal model

Two types of estimation can be conducted using the Spatiotemporal model. The first one is the estimation of the fixed effects jointly with the random effects, referred to as the calibration, and the second one is the estimation of the random effects of a patient knowing the fixed effects, named personalisation. An implementation of the Spatiotemporal model enabling both estimations is available in the open-source package Leaspy: `https://gitlab.com/icm-institute/aramislab/leaspy`.

2.2.5.1 Model calibration

The first step is the calibration of the model which enables the estimation of the fixed and associated random effects from a training dataset. There is no analytical solution for maximising the log-likelihood. Thus we used an Expectation-Maximization algorithm. Nevertheless, the computation of the expectation is also intractable due to the non-linearity of the model. Thus, a Monte-Carlo Markov Chain Stochastic Approximation Expectation-Maximization

(MCMC-SAEM) algorithm (Kuhn and Lavielle, 2004) was used with a Robbins-Monro convergence algorithm (Robbins and Monro, 1951) applied on the last iterations to get the mean of the distribution of the model. Note that convergence of the MCMC-SAEM algorithm has been proven for models that lie in the curved exponential family (Kuhn and Lavielle, 2004), a category in which falls the Spatiotemporal model. More details are given by Koval, 2020 (p.41-43) and Schiratti, 2017 (p.106).

2.2.5.2 Model personalisation

The second step is personalisation. It is mainly used to compute the random effects for new patients from a test dataset. During this step, we used previously computed fixed effects from the calibration. During this personalisation, only the random effects are estimated, to get the individual longitudinal processes. The solver *minimise* from the package Scipy (Virtanen et al., 2020) was used to maximise the log-likelihood.

2.3 Method: Analysis of ALS heterogeneity

2.3.1 Patients

We conducted our analysis using data from the Pooled Resource Open-Access ALS Clinical Trials Consortium (PRO-ACT) database. This database is a compilation of 23 phase II and III clinical trials along with one observational study. Notably, the database does not include any information that could potentially lead to patient identification, such as the clinical trial, tested drug, study centres, or dates. Given that the database is an aggregation of various sources, it encompasses multiple inclusion and exclusion criteria for patients entering the cohort. More detailed information can be found in the paper that introduces the database (Atassi et al., 2014).

We selected patients with a minimum of two visits recorded for all the outcomes listed below. We collected the following data: values of the outcomes at each available visit with the age of the patient at the respective visits, patient sex, and onset site (spinal or bulbar), a label available within the PRO-ACT database. Patients were categorized into four subgroups based on sex and onset site: men with bulbar onset, women with bulbar onset, men with spinal onset, and women with spinal onset.

2.3.2 Outcomes

We selected a set of outcomes that are important clinical proxies of ALS progression and were available in the PRO-ACT database. We considered the most widely used functional rating system in patients with ALS, namely the revised version of the ALS functional rating scale (ALSFRSr)(Rooney et al., 2017). The scale starts at a maximum theoretical value of 48 and decreases with the severity of the disease till zero. We considered the four ALSFRSr sub-scores assessing four distinct domains: bulbar, gross motor, fine motor, and respiratory functions. Each sub-scale has 12 points each (Rooney et al., 2017). In addition, we considered the BMI (Moglia et al., 2019) and the Forced Vital Capacity (FVC)(Daghlas,

Govindarajan, and Pooled Resource Open-Access ALS Clinical Trials Consortium, 2021) that we normalized using the computed normal FVC (Brändli et al., 1996).

We normalized the outcomes between 0 (the healthiest value) and +1 (the maximum pathological change). The four ALSFRSr sub-scores were normalized using their theoretical maximum and minimum values. For FVC and BMI, we first applied a box-cox transformation (estimated on the whole PRO-ACT dataset) to get a standard normal distribution (Pedregosa et al., n.d.; Yeo and Johnson, 2000; Box and Cox, 1964). We then clipped the resulting value between -3 and 3 points and converted them to the 0 to 1 scale.

In addition to these repeated measures, we collected the following covariates when available: riluzole use, trial arm (active, placebo, observational), age at first symptom, age at diagnosis, age at baseline, Mitos (Fang et al., 2017) and FT9 scores (Thakore et al., 2018) at baseline. These data were not used as input for the disease progression model.

2.3.3 Statistical Analysis

We calibrated the model with the dataset described above and composed of four subgroups for 20,000 iterations.

The four studied subgroups had different sizes, so we wanted to assess if the model calibrated all of them well. To do so, we computed the reconstruction error as the difference between the estimated value of each visit and its real available value and compared the reconstruction error between subgroups.

We used ANOVAs to test for the interaction of sex and onset site on the random effects of the Spatiotemporal model. When the results were significant, we conducted independent pairwise Tukey-Kramer tests (Tukey, 1949) for further comparison. Conversely, if no significant interaction was observed, we conducted an ANOVA without interaction to assess the influence of each covariate while adjusting for others. All the results are presented with a 95% confidence interval computed with pairwise Tukey-Kramer or ANOVA without interaction, depending on the interaction significance.

The same comparison process was used to compare the subgroups at baseline and describe the database. For database description, data are expressed as mean +/- standard deviation or frequency (percent). Statistical analysis used a conventional 2-tailed type I error of 0.05. Data were analyzed using packages statsmodel (Skipper and Perktold, 2010) and bioinfokit (Bedre, 2020) for Python.

2.4 Results

2.4.1 Patients

In the PRO-ACT database, which included 8,571 patients (with a total of 78,824 visits), we excluded 220 BMI values and 45 FVC values that fell outside the normalization range. Subsequently, we selected the 1,463 patients with at least two visits for all the selected outcomes. Finally, we further refined our selection to 1,438 patients who had information

regarding the onset site. Note that sex was available for all patients. Further characteristics of the population are presented in Figure 2.1.

The database consisted of 51% men with spinal onset, 12% men with bulbar onset, 26% women with spinal onset, and 11% women with bulbar onset. Over the four subgroups, there were no significant differences in follow-up duration (1.00 ± 0.59 years), the number of visits per patient (11.6 ± 7.2 visits), and time between visits (34.5 ± 28.5 days). However, significant differences were observed among subgroups concerning ages, disease duration, and some baseline scores.

Outcomes			Spinal Man	Bulbar Man	Spinal Woman	Bulbar Woman	p-value	Subgroup relations
Numbers	Patients		757	126	405	150		
	Visits		9236	1447	4771	1799		
	Visits per patient		12.2 ± 10.0	11.5 ± 7.4	11.8 ± 9.0	12.0 ± 7.7	0.46	-
Riluzole use	Took at least once		533 (70%)	82 (65%)	286 (71%)	96 (64%)		
	Never took		126 (17%)	31 (25%)	72 (18%)	29 (19%)		
	No information on use		98 (13%)	13 (10%)	47 (12%)	25 (17%)		
Trial arm	Active		378 (50%)	66 (52%)	204 (50%)	65 (43%)		
	Placebo		210 (28%)	35 (28%)	115 (28%)	40 (27%)		
	Observational		169 (22%)	25 (20%)	86 (21%)	45 (30%)		
Age	First symptoms		52.3 ± 11.1 (100%)	53.1 ± 11.6 (100%)	54.1 ± 10.9 (100%)	59.5 ± 9.5 (100%)	0.0021	(SM,BM,SW)<BW
	Diagnosis		53.0 ± 11.1 (87%)	53.7 ± 11.7 (89%)	54.7 ± 10.6 (85%)	60.5 ± 9.5 (90%)	0.0012	(SM,BM,SW)<BW
	Baseline		54.1 ± 11.1 (100%)	54.6 ± 11.7 (100%)	56.0 ± 10.8 (100%)	61.1 ± 9.4 (100%)	0.0018	(SM,BM,SW)<BW
Time duration	Patient time of follow-up (y)		1.0 ± 0.7 (100%)	1.0 ± 0.6 (100%)	1.0 ± 0.8 (100%)	1.0 ± 0.6 (100%)	0.58	-
	Time between visits (d)		34 ± 28 (100%)	35 ± 28 (100%)	35 ± 30 (100%)	34 ± 29 (100%)	0.33	-
	Disease duration (y)		1.8 ± 1.0 (100%)	1.4 ± 0.9 (100%)	1.9 ± 1.1 (100%)	1.6 ± 0.9 (100%)	0.61	B<S***
Outcomes at baseline	Mitos		0.2 ± 0.5 (59%)	0.2 ± 0.5 (59%)	0.3 ± 0.5 (61%)	0.2 ± 0.5 (71%)	0.70	-
	FT9		1.8 ± 0.9 (59%)	1.9 ± 1.0 (59%)	1.9 ± 0.9 (61%)	1.9 ± 1.0 (71%)	0.52	-
	ALSFRSr score	Total	37.2 ± 5.4 (59%)	37.4 ± 5.8 (59%)	35.9 ± 5.5 (61%)	37.3 ± 5.3 (71%)	0.19	W<M***
		Respiratory	11.3 ± 1.3 (59%)	11.1 ± 1.3 (59%)	11.2 ± 1.3 (61%)	11.1 ± 1.2 (71%)	0.73	-
		Bulbar	10.8 ± 1.5 (59%)	7.6 ± 2.1 (59%)	10.5 ± 1.7 (61%)	6.7 ± 2.3 (71%)	0.025	BW<BM<(SW,SM)
		Fine motor	7.6 ± 2.8 (59%)	9.2 ± 2.6 (59%)	8.0 ± 2.7 (61%)	10.0 ± 2.1 (71%)	0.39	M<W**;B<S***
		Gross motor	7.6 ± 2.5 (59%)	9.5 ± 2.5 (59%)	6.2 ± 2.4 (61%)	9.5 ± 2.3 (71%)	0.0012	SW<SM<(BW,BM)
	Forced vital capacity		75.9 ± 14.4 (97%)	70.3 ± 13.6 (96%)	74.6 ± 14.4 (96%)	68.7 ± 13.8 (97%)	0.83	B<S***
	BMI		27.1 ± 4.2 (98%)	26.8 ± 4.3 (97%)	26.4 ± 5.0 (98%)	25.6 ± 4.9 (98%)	0.46	W<M***

TABLE 2.1: Patient characteristics

*Data are expressed as mean± standard deviation (percentage of available data) or frequency (percentage), p-value: ANOVA interaction p-value, subgroup relations: if the interaction was significant, significant relations between the subgroups are described, otherwise only the relation between sex and onset without interaction is described. BM: bulbar men, BW: bulbar women, SM: spinal men, SW: spinal women, *: <0.05, **: <0.01, ***: <0.001*

2.4.2 Reconstruction error

The maximum differences of the median (orange line in Figure 2.3) were small compared to the observed median of patient follow-up variation (difference between the minimum and maximal value of the outcome across patient follow-up) for each outcome considered: 2.18% compared to 17.35% for FVC, 0.05 point compared to 2.50 kg/m^2 for BMI, 0.19 point compared to 2.00 point for respiratory scale, 0.31 point compared to 2.00 point for bulbar scale, 0.08 point compared to 4.00 point for fine motor scale and 0.07 point compared to 4.00 point for gross motor scale.

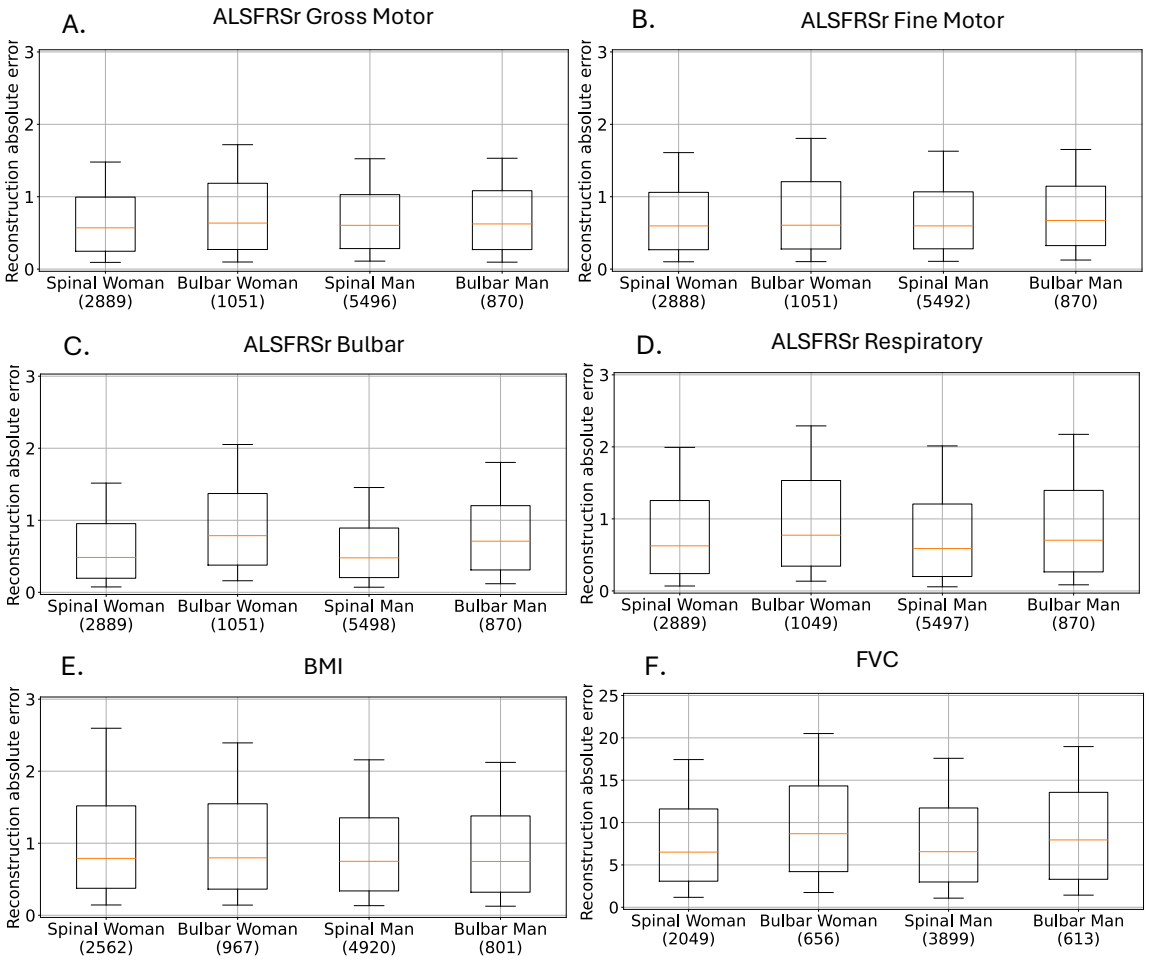


FIGURE 2.3: Reconstruction error distribution of the Longitudinal Spatiotemporal model per outcome and subgroup

Box plots are represented with the median in orange in the native scale of the outcome. The numbers in parentheses represent the number of visits on which the boxplot was computed.

2.4.3 ALS outcomes trajectories

The average disease progression was first extracted for the four subgroups. Figure 2.4 represents the progression of ALSFRSr subscores for each subgroup. Overall, curves are shifted on the right for patients with bulbar onset (Figure 2.4 B and D) compared to patients with spinal onset (Figure 2.4 A and C). Curves are also shifted on the right for

women (Figure 2.4 A and B) compared to men (Figure 2.4 C and D), illustrating later onset for women compared to men.

The model also exhibited differences in the order of progression of the different outcomes. For instance, it confirmed the progression order of ALSFRSr bulbar curve (orange) and ALSFRSr motor curves (red and green): they are switched between patients with spinal onset (Figure 2.4 A and C) and patients with bulbar onset (Figure 2.4 B and D). Finally, we can note differences in speed. Patients with spinal onset (Figure 2.4 A and C and Figure 2.5 purple and grey) seem to have steeper curves than patients with bulbar onset (Figure 2.4 B and D and Figure 2.5 cyan and pink). All these results are described below by quantitative statistical tests.

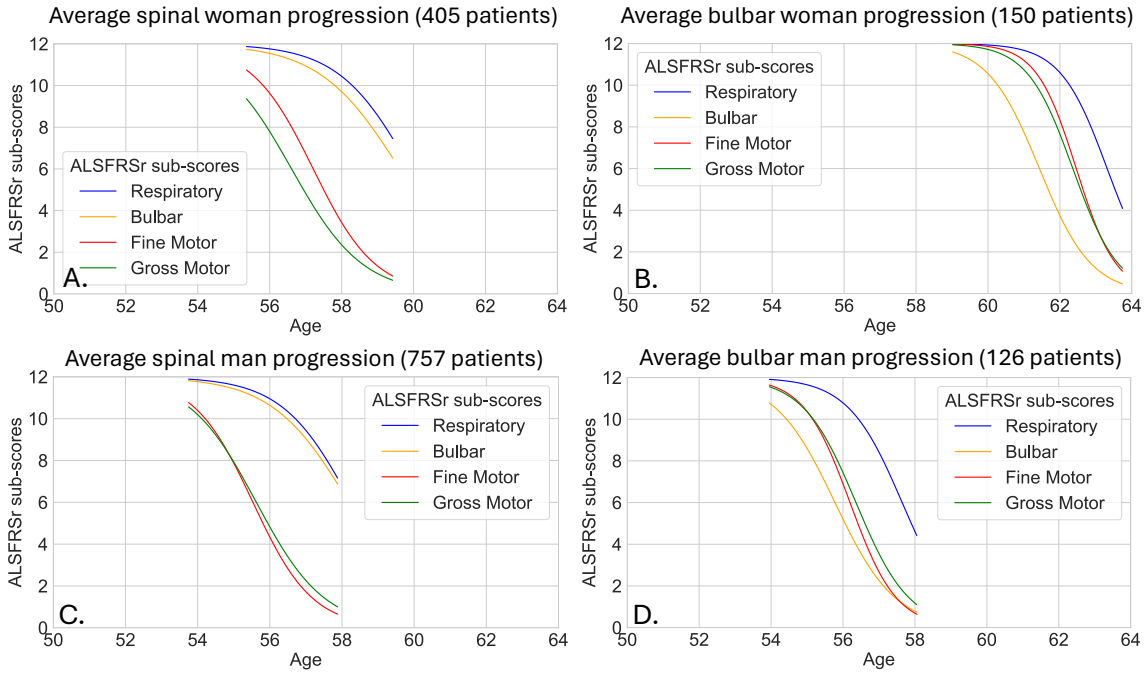


FIGURE 2.4: Average progression of ALSFRSr subscores depending on sex and onset site *Respiratory (blue), Bulbar (orange), Fine Motor (red), Gross Motor (green)*

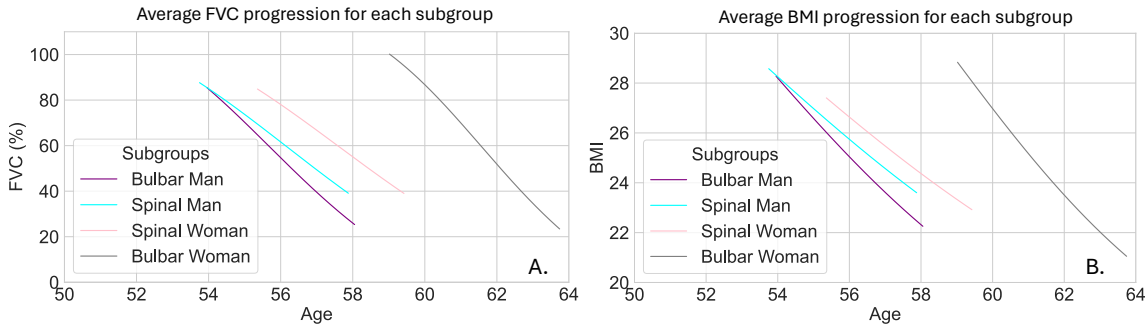


FIGURE 2.5: Average subgroup progressions of Forced Vital Capacity (FVC) and Body Mass Index (BMI) for each subgroup

2.4.4 Individual temporal variability (ξ_i, τ_i)

2.4.4.1 Estimated age at disease onset (τ_i)

We found an interaction between the onset site and sex for the reference time (p = 0.0020) (Figure 2.6 A). For patients with bulbar onset, the reference time for men occurred 6.0 years earlier (95% CI = [2.6, 9.5]) than for women while for a woman with spinal onset, it occurred 4.7 years earlier (95% CI = [1.9, 7.4]) than for women with bulbar onset.

2.4.4.2 Speed of progression (ξ_i)

We did not find any significant interaction between the onset site and sex for the speed factor of progression (p-value = 0.14) (Figure 2.6 B). Patients with bulbar onset were found to progress 1.4 times faster (95% CI = [1.2, 1.6]) than patients with spinal onset independently of sex.

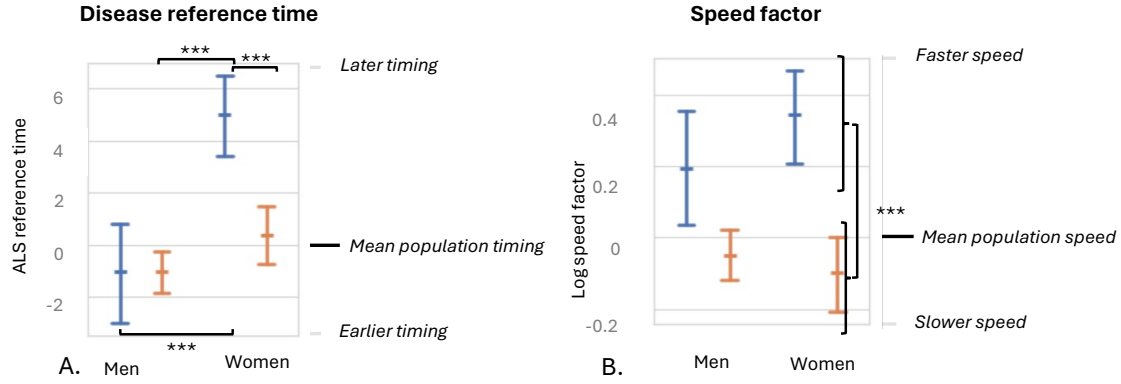


FIGURE 2.6: ALS estimated speed factor and reference time depending on sex and onset site

*Panel A: the vertical axis presents the estimated reference time in months compared to the mean estimated reference time of the whole population. Panel B: The vertical axis presents the log speed compared to the mean log speed of the whole population. Both horizontal axes are for sex with estimation for the bulbar patients in blue and the spinal patients in orange. The graph displays the mean of this delay with its confidence interval.ANOVA interaction p-value is equal to: (A) 0.002 for the reference time and (B) 0.14 for the log speed factor . ***: p-value for pairwise comparison <0.001*

2.4.5 Individual spatial variability ($w_i v_0^{-1}$)

Space-shifts (w_i) were corrected by the speed (v_0) to facilitate there interpretation in a time unit.

2.4.5.1 Motor decline

We found an interaction between sex and onset site for the gross motor scale (p<0.001) but not for the fine motor scale (p-value = 0.12), once corrected for the estimated reference timing and speed of progression (Figure 2.7 A and B). Among patients with spinal onset, ALSFRSr gross motor scale deteriorated 4.5 months later (95% CI = [3.1, 6.0]) in women

than in men. However, ALSFRSr fine motor scale changed 3.5 months earlier (95% CI = [2.6, 4.4]) in women than in men, independently of the onset site. Among men, ALSFRSr gross motor scale changed 10.6 months earlier (95% CI = [8.3, 12.8]) for patients with spinal onset compared to bulbar onset. Among women, the difference increased to 16.0 months earlier (95% CI = [13.7, 18.2]). ALSFRSr fine motor scale changed 10.2 months earlier (95% CI = [9.0, 11.3]) for patients with spinal onset compared to bulbar onset, independently of the sex.

2.4.5.2 Bulbar signs decline

We observed an interaction between sex and onset site for ALSFRSr bulbar score (p-value = 0.01), once corrected for the estimated reference timing and speed of progression (Figure 2.7 C). Among patients with bulbar onset, ALSFRSr bulbar scale changed 6.4 months earlier (95% CI = [1.6, 11.1]) in women than men. Among men, the ALSFRSr bulbar scale changed 27.2 months later (95% CI = [23.4, 31.0]) for patients with spinal onset compared to bulbar onset. Among women, the difference increased to 32.5 months later (95% CI = [28.7, 36.3]).

2.4.5.3 Respiratory decline

We found an interaction between sex and onset site for FVC ($p<0.001$) but not for ALSFRSr respiratory score (p-value = 0.058), once corrected for the estimated reference timing and speed of progression (Figure 2.7 E and F). Among patients with bulbar onset, FVC was impaired 2.6 months earlier (95% CI = [0.6, 4.6]) in women than in men. Among men, FVC was impaired 5.4 months earlier (95% CI = [3.8, 7.0]) in patients with bulbar onset compared to spinal onset. Among women, the difference increased to 8.3 months (95% CI = [6.7, 9.9]). ALSFRSr respiratory scale changed 2.4 months earlier (95% CI = [0.7, 4.0]) for patients with bulbar onset compared to spinal onset independently of the sex.

2.4.5.4 BMI decline

We did not observe an interaction between sex and onset site for BMI (p-value = 0.17), once corrected for the estimated reference timing and speed of progression (Figure 2.7 D). BMI decreased 8.9 months earlier (95% CI = [3.9, 13.8]) in women than men, independently of the onset site. BMI decreased 10.0 months earlier (95% CI = [3.8, 16.1]) for patients with bulbar onset compared to spinal onset, independently of the sex.

2.5 Discussion

2.5.1 Results interpretation

2.5.1.1 Methodology

The reconstruction error of the model was of the same range as the one of the Alzheimer's disease study (Koval et al., 2021), for which the model was originally developed, which confirmed the potential of the Spatiotemporal model in ALS. The strength of the Spatiotemporal model is its ability to synchronise patients onto a common disease timeline

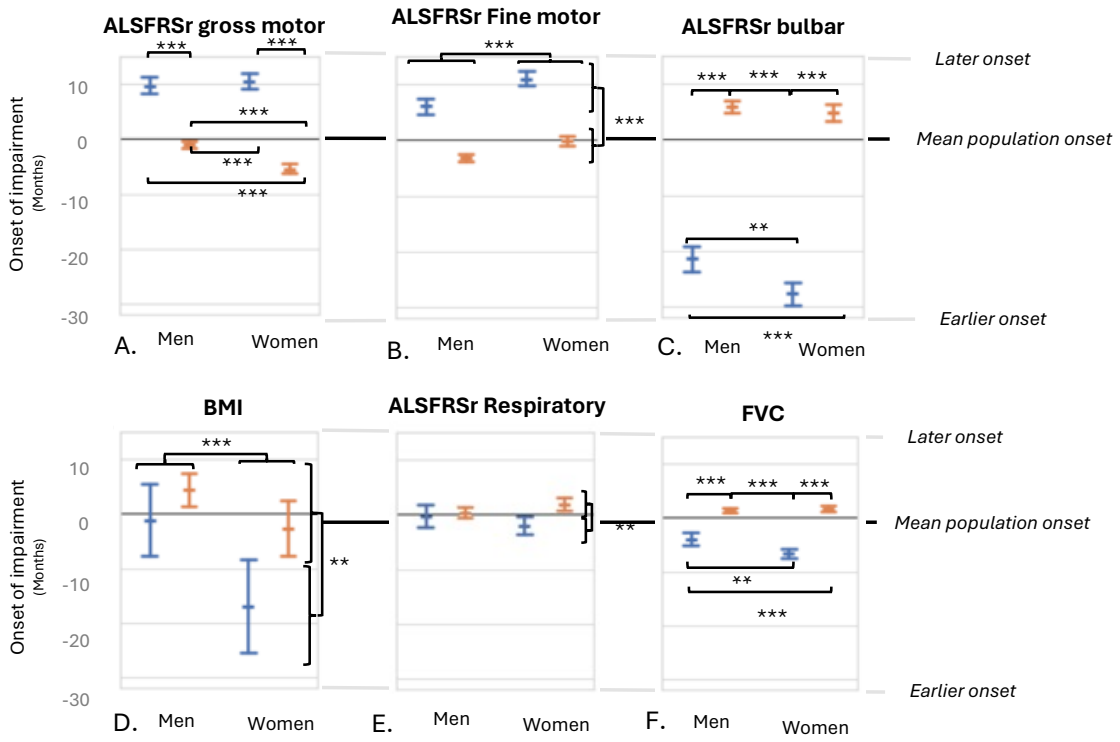


FIGURE 2.7: Onset of impairment depending on sex and onset site
*The vertical axis presents the delay of outcome impairment onset in months compared to the mean onset of the whole population. The horizontal axis is for sex with estimation for the bulbar patients in blue and the spinal patients in orange. The graph displays the mean of this delay with its confidence interval. Results are corrected for the subgroup estimated reference timing and speed of progression. ANOVA interaction p-value is (A,C,D) below 0.001 for ALSFRSr gross motor, ALSFRSr bulbar, and FVC, (B) 0.12 for ALSFRSr fine motor, (E) equal to 0.058 for ALSFRSr respiratory, and (D) 0.17 for BMI. FVC: Forced Vital Capacity (%), **: p-value for pairwise comparison <0.01 ,***: p-value for pairwise comparison <0.001*

(temporal aspect) thanks to a latent disease age, while also capturing the remaining variability through parameters that account for outcome ordering (spatial aspect). Thanks to its interpretable structure, random effects distribution could be analysed to study ALS heterogeneity in terms of speed of progression, estimated age at onset and order of outcomes decline.

2.5.1.2 Clinical

Concerning the effect on weight, we showed that BMI starts to decrease 9 months earlier in women than men, independently of the onset site. Considering the rapid progression of ALS, with a mean survival of three to four years (Talbott, Malek, and Lacomis, 2016), this inter-sex difference appears clinically meaningful. Weight loss is a classical consequence of bulbar impairment leading to dysphagia. However, alterations in body weight are present in ALS patients already decades before the clinical manifestation of ALS (Peter et al., 2017) and are a negative prognostic factor associated with shorter survival (Li et al., 2022). Weight loss is also associated with increased energy expenditure. It is present in almost half

of the patients and is associated with shorter survival (Nakamura et al., 2022). Inter-sex differences in metabolism may also affect the disease trajectory. The role of energy and lipid metabolism abnormalities in the pathogenesis of ALS is currently a focus of many studies in animal models and ALS patients (Dupuis et al., 2011; Nelson and Trotti, 2022; Vandoorne, De Bock, and Van Den Bosch, 2018; Guillot, Bolborea, and Dupuis, 2021). Inter-sex differences in the percentage of body fat may also modify disease progression. It is higher in women (Blaak, 2001) and may vary with menopausal status (Toth et al., 2000). Studies comparing fat mass with impedancemetry or dual-energy X-ray absorptiometry are needed to investigate better the role of metabolism and the effect of sex on disease progression (Lee et al., 2021).

The effect of sex on respiratory decline has not been reported so far. We showed that among patients with bulbar onset, FVC is impaired 2.6 months earlier in women than in men. This difference is also clinically meaningful since respiratory distress is the main cause of death in ALS and the decline of FVC is the main criterion for the indication of non-invasive ventilation (Morelot-Panzini, Bruneteau, and Gonzalez-Bermejo, 2019).

An unexpected result was that the ALSFRSr fine motor scale changes earlier in women, regardless of the onset site, but that the ALSFRSr gross motor scale changes later in women among patients with spinal onset. Further studies are needed to investigate whether these differences might be related to sexual dimorphism in the anatomic pattern of abnormalities, both in terms of susceptibility to neurodegeneration or compensatory mechanisms. The advances in neuroimaging enable an increasing number of research studies investigating sexual dimorphism in both health and disease conditions (Bede et al., 2022; Trojsi et al., 2021). In ALS more specifically, significant sex differences in the anatomical patterns of cortical and subcortical pathology have been shown (Bede et al., 2022). Finally, these sex differences might be a bias in the self-evaluation of motor abilities using ALSFRSr.

From a biological perspective, the mechanisms of sexual dimorphism in ALS have been related to sex-specific microglial regulation (Lay and Li, 2019), sex hormones (Raymond et al., 2021), or susceptibility to oxidative stress (Sumien et al., 2021). A study, performed on a large number of ALS women with natural menopause and well-defined oral contraceptive usage, has demonstrated that longer exposure to female hormones has neuroprotective effects on motoneurons in ALS (positive association between a longer reproductive condition, the susceptibility to ALS, and the survival of ALS patients)(Raymond et al., 2021).

2.5.2 Strengths & Weaknesses

2.5.2.1 Methodological

We used a mixed-effect model following neurological research recommendations (Locascio and Atri, 2011). Logistic curves were used to overcome common limitations on score analysis: curvilinearity, and potential floor or ceiling effects (Gordon et al., 2010). A population-based study showed the existence of plateaux in ALSFRSr mainly shorter than 6 months, with only 10% of patients with wider plateaux (Vasta et al., 2020). Thus, modelling the progression

of ALSFRSr subscores by logistic curves could be debated. Nevertheless, plateaux appear in high scores and are unlikely to change the results since our patients' average length of follow-up is one year.

The Spatiotemporal model used is robust to missing data and does not require a standardized time between visits (Couronne et al., 2019). Even though, the informative right censoring of the visits by death might introduce biases in our study (Schluchter, 1992). Thus, studying jointly longitudinal and survival data, potentially with a joint model (see section 1.4) can result in more precise estimates and improve inference (Lu et al., 2023). Using a joint model could also bring insight into the interaction between sex and onset site corrected by the progression of the longitudinal outcomes.

2.5.2.2 Clinical

A strength of our study is that our model was based on a large, multicentric, and longitudinal database. However, the PRO-ACT database may not represent the real-life ALS patient population. Since the database aggregates clinical trial data, there is an over-representation of men, young subjects, patients with spinal onset, and slow progressors. Nevertheless, these results could help better understand the progression of placebo arms and design strategies for patient stratification in clinical trials. The absence of information on genetic mutation (Swinnen and Robberecht, 2014), neuropsychological symptoms (Swinnen and Robberecht, 2014), and lifestyle (Trojsi et al., 2012; Westeneng et al., 2021) also limits the interpretation of the results.

2.6 Conclusion & perspectives

To summarize, we showed that sex and onset site interaction is an important driver of ALS progression considering many clinical proxies such as disability, weight loss, and respiratory function. Pathophysiological mechanisms underlying these differences are elusive and may account for an interplay of biological, lifestyle, and environmental factors.

From a clinical perspective, further studies in animal models and patients are needed to understand better their respective contributions (Chiò et al., 2020). In addition, from a methodological point of view, this study confirms the relevance of the Spatiotemporal model in ALS and calls for a joint Spatiotemporal model.

Chapter 3

A univariate Joint Temporal model

In this chapter, we propose a univariate Joint Temporal model, that would enable to jointly model longitudinal and survival data while automatically realigning partial patient trajectories. To do so, we used the longitudinal Spatiotemporal model with its latent disease age, as a longitudinal submodel. We associated it with a survival submodel that estimates a Weibull distribution from the latent disease age. First, we validated our model on different simulated scenarios. Then, we benchmarked our model with a state-of-the-art joint model and reference survival and longitudinal models on simulated and real data in the context of ALS. On PRO-ACT ALS data, our model got significantly better results than the state-of-the-art joint model for absolute bias (4.21(4.41) versus 4.24(4.14)(p-value=1.4e-17)), and mean cumulative AUC for right censored events (0.67(0.07) versus 0.61(0.09)(p-value=1.7e-03)). Thus, we showed that our approach is better suited than the state-of-the-art in the context where the reference time to realign trajectories is not reliable.

This chapter has been submitted. See (Ortholand et al., 2024).

Contents

3.1 Introduction

As described in section 1.2.1, longitudinal data available for progressive chronic diseases are often sparse and cover only parts of the progression, which emphasises the need to realign manually or automatically patients' ages depending on their disease stages, to extract a full typical timeline of the disease (Young et al., 2024). The date of the first symptom is often used to do so, but in such chronic disease, it might not be a reliable reference time as onset is suspected to occur years before clinical manifestations (Jack et al., 2010; Peter et al., 2017). In the case of ALS, survival data (see section 1.3.1), with the occurrence of death, are also of importance notably to demonstrate a treatment effect (FDA, 2019).

Joint models enable modelling both survival and longitudinal data together, which can enable to reduce biases when they are associated with the same biological processes (Lu et al., 2023). Joint models are composed of three parts: a model for the survival data, a model for the longitudinal data and a linkage structure, which captures the link between the two types of data (Rizopoulos, 2012; Elashoff, Li, and Li, 2016; Hickey et al., 2016; Furgal, Sen, and Taylor, 2019; Alsefri et al., 2020).

The most used survival model, for data with right non-informative censoring, is the Cox Proportional Hazard (Cox) model (Cox, 1972), with a proportional effect of covariates on the hazard scale (see section 1.3.2). Nevertheless, as the baseline hazard needs to be estimated for joint models (Rizopoulos, 2012) and the proportional assumption is often violated on long follow-up, it might not be the appropriate survival submodel. Thus Accelerated Failure Time (AFT) models (Kalbfleisch and Prentice, 2002) may be better suited as survival submodels.

As described in section 1.4, joint models are built upon Generalised Linear Mixed-effect Models (GLMMs) (McCulloch, Searle, and Neuhaus, 2008) as longitudinal submodel. But such classical mixed-effect models rely on an empirical disease time axis, which limits their temporal resolution to the resolution of the reference time used to index the disease time axis (Schiratti et al., 2015; Schiratti, 2017; Young et al., 2024). To overcome this issue, longitudinal models that capture the data-driven disease timeline behind the observable data, named data-driven disease progression models, were developed (Young et al., 2024) as described in section 1.2.3. Among others, the Spatiotemporal model was proposed to capture, from a population sample, a latent disease age to overcome the need for a reference time (Schiratti et al., 2015; Schiratti, 2017).

As mentioned in section 1.4, three main types of joint modelling were proposed: the latent class model (Proust-Lima, Philipps, and Liquet, 2017; Proust-Lima et al., 2023), the frailty joint model (Rondeau, Mazroui, and Gonzalez, 2012) and the shared random effects model (Rizopoulos, 2016). Compared to the other two types of models, the advantage of the shared random effects model is that it captures individual variability without the need for correlated survival data. Nevertheless, the shared random effects models have their own limitations: by including predictors of the longitudinal outcome in the survival model, they usually focus on how longitudinal outcomes impact survival. They also rely on GLMM which limits their time resolution. Modelling ALS progression raises thus the need for a Joint data-driven model.

From a clinical point of view, having a personalised prognosis in ALS could help, in addition to better-anticipating care management of the patient, to design more powered trials (Maheux et al., 2023; Raket, 2022). An online tool is freely available online, for neurologists, to give insight on prognosis (Westeneng et al., 2018)(http://www.encalssurvivalmodel.org/) and it was also the subject of the challenge Prize4Life (Kueffner et al., 2019).

In this chapter, we propose a Joint Temporal model suited for progressive chronic diseases that overcomes the need, of the state-of-the-art joint models, for a reliable reference time. To do so, we used the longitudinal Spatiotemporal model as the longitudinal submodel and used its defined latent disease age as the linkage structure. We associated a survival submodel that estimates a Weibull distribution from the latent disease age, in the spirit of an AFT model.

After introducing the Joint Temporal model, we validated it using sensitivity analysis on simulated data corresponding to different real-like clinical scenarios. We then benchmarked the Joint Temporal model against reference models on simulated and real ALS data, and showed that the proposed approach is better suited than the state-of-the-art when reference time used to realign patients is not reliable.

3.2 Model specifications

3.2.1 Notations

3.2.1.1 Data

As in section 2.2.1, we consider N patients, indexed by i, followed for n_i visits indexed by j and observed at an age $t_{i,j}$. In this chapter, only one outcome $y_{i,j}$ is measured. We assume that we also observe an event e, and denote t_{e_i} the age of the patient when the event is observed. To distinguish censored and observed events, a boolean B_{e_i} is associated with the time of the event t_{e_i}: $B_{e_i} = 0$ if the event is censored and $B_{e_i} = 1$ if the event is observed (see section 1.3.1). If the event is censored, the time of the last visit is used as the time of the event t_{e_i} (Leung, Elashoff, and Afifi, 1997).

3.2.1.2 Joint model structure

The objective of joint models is to describe the combination of two types of clinical data: longitudinal data and survival data, with their linkage. As detailed in section 2.2.3.1, the longitudinal process, is modelled by $\gamma_i(t_{i,j})$ so that $y_{i,j} = \gamma_i(t_{i,j}) + \epsilon_{i,j}$ with $\epsilon_{i,j} \sim \mathcal{N}(0, \sigma)$ a Gaussian noise. The survival process $S_i(t)$, is modelled by the probability that a patient i experiences the event after age t with $S_i(t) = p(t_{e_i} > t)$.

3.2.2 The Joint Temporal model

The Joint Temporal model is built upon a univariate version of the Spatiotemporal model described in section 2.2.3. In such a univariate model, only temporal variability is allowed and no random effects are capturing spatial variability (see section 2.2.3.2) as there is only one outcome y considered.

3.2.2.1 The latent disease age: correction of individual variation

In the Joint Temporal model, the differences in the observed univariate trajectories are encapsulated in the latent disease age $\psi_i(t) = e^{\xi_i}(t - \tau_i) + t_0$ with e^{ξ_i} the individual speed factor and τ_i the individual estimated reference time (see section 2.2.3.2). The latent disease age is then used as the link between the longitudinal and the survival processes $(\gamma_i(t), S_i(t))$, both estimated from this latent disease age.

$$\begin{cases} \gamma_i(t) = \gamma_0(\psi_i(t)) \\ S_i(t) = S_0(\psi_i(t)) \end{cases}$$

Indeed, for the survival modelling, instead of using time 0 as a start time, we used the reference time t_0 and imposed $\forall t < t_0, S_0(t) = 1$. This assumption makes sense as t_0 enables realigning trajectories and corresponds to a time of a given value of the score that most of the patients experimented with.

3.2.2.2 Modelling longitudinal process

Following what was described in section 2.2.3.3, the univariate Spatiotemporal model can be rewritten for a patient i and a visit j as:

$$\gamma_i(t_{i,j}) = \left(1 + g \times \exp(-v_0 \tfrac{(g+1)^2}{g}(\psi_i(t_{i,j}) - t_0))\right)^{-1}$$

3.2.2.3 Modelling survival process

The Weibull distribution is used to model the survival probability from the latent disease age $\psi_i(t)$:

$$\begin{aligned} S_i(t) &= S_0(\psi_i(t)) \\ &= \mathbb{1}_{\psi_i(t)>t_0} \exp\left(-\left(\tfrac{\psi_i(t)-t_0}{\nu}\right)^{\rho}\right) + \mathbb{1}_{\psi_i(t)\leq t_0} \\ &= \mathbb{1}_{t>\tau_i} \exp\left(-\left(\tfrac{e^{\xi_i}(t-\tau_i)}{\nu}\right)^{\rho}\right) + \mathbb{1}_{t\ \leq \tau_i} \end{aligned}$$

where ν represents the scale and ρ the shape of the Weibull distribution. From there we also compute the individual hazard, which is, assuming that a patient has survived for a time t, the probability that he will not survive for an additional time dt:

$$\begin{aligned} h_i(t) &= -\frac{S_i'(t)}{S_i(t)} \\ &= \mathbb{1}_{\psi_i(t)>t_0} \tfrac{\rho e^{\xi_i}}{\nu} \left(\tfrac{\psi_i(t)-t_0}{\nu}\right)^{\rho-1} \\ &= \mathbb{1}_{t>\tau_i} \tfrac{\rho e^{\xi_i}}{\nu} \left(\tfrac{e^{\xi_i}(t-\tau_i)}{\nu}\right)^{\rho-1} \end{aligned}$$

3.2.2.4 Joint Temporal model

The Joint Temporal model is the combination of both a longitudinal submodel $\gamma_i(t)$ and a survival submodel $S_i(t)$ using the latent disease age $\psi_i(t)$ as a linkage structure:

$$\begin{cases} \psi_i(t) = e^{\xi_i}(t - \tau_i) + t_0 \\ \gamma_i(t) = \gamma_0(\psi_i(t)) = \left(1 + g \times \exp(-v_0 \frac{(g+1)^2}{g}(\psi_i(t) - t_0))\right)^{-1} \\ S_i(t) = S_0(\psi_i(t)) = \mathbb{1}_{\psi_i(t)>t_0} \exp\left(-\left(\frac{\psi_i(t)-t_0}{\nu}\right)^{\rho}\right) + \mathbb{1}_{\psi_i(t)\leq t_0} \end{cases} \tag{3.1}$$

3.2.3 Likelihood of the Joint Temporal model

3.2.3.1 Parameters

We put ourselves in the same context as in section 2.2.4.1. The parameters of the model can be summarised as follows for each patient i:

- Latent parameters (z):
  - Latent fixed effects (z_{fe}): fixed effects sampled

$$\tilde{g} = \log(g) \sim \mathcal{N}\left(\overline{\tilde{g}}, \sigma^2_{\tilde{g}}\right) \qquad \tilde{v_0} = \log(v_0) \sim \mathcal{N}\left(\overline{\tilde{v_0}}, \sigma^2_{\tilde{v_0}}\right)$$
$$\tilde{\nu} = -\log(\nu) \sim \mathcal{N}\left(\overline{\tilde{\nu}}, \sigma^2_{\tilde{\nu}}\right) \qquad \tilde{\rho} = \log(\rho) \sim \mathcal{N}\left(\overline{\tilde{\rho}}, \sigma^2_{\tilde{\rho}}\right)$$

  - Latent random effects (z_{re}): random effects sampled

$$\xi_i \sim \mathcal{N}\left(\overline{\xi}, \sigma^2_{\xi}\right) \qquad \tau_i \sim \mathcal{N}\left(\overline{\tau}, \sigma^2_{\tau}\right)$$

- Model parameters (θ): fixed effects estimated from log-likelihood maximisation $\theta = \{\sigma_\xi, \sigma_\tau, t_0, \overline{\tilde{g}}, \overline{\tilde{v_0}}, \overline{\tilde{\nu}}, \overline{\tilde{\rho}}, \sigma\}$
- Hyperparameters (Π): set by the user $\Pi = \{\sigma_{\tilde{g}}, \sigma_{\tilde{v_0}}, \sigma_{\tilde{\nu}}, \sigma_{\tilde{\rho}}\}$

To ensure identifiability, we set $\overline{\xi} = 0$ and $t_0 = \overline{\tau}$.

3.2.3.2 Log-likelihood

The likelihood estimated by the model is the following:

$$p(y, T_e, B_e \mid \theta, \Pi) = \int_z p(y, T_e, B_e, z \mid \theta, \Pi) dz$$

$p(y, T_e, B_e, z \mid \theta, \Pi)$ can be divided into two different terms: data attachment which represents how well the model describes the data (y, t_e, B_e) and a prior attachment, which prevents over-fitting.

$$\log p((y, t_e, B_e), z \mid \theta, \Pi) = \quad \log p(y, t_e, B_e \mid z, \theta, \Pi) + \log p(z, \mid \theta, \Pi)$$

The first term, data attachment, can be divided again into two terms considering that survival and longitudinal processes are independent knowing the random effects, which is an

assumption common to other joint models (Rizopoulos, 2012; Proust-Lima, Philipps, and Liquet, 2017). We can also split the prior attachment term: two terms for the prior attachment of latent parameters (fixed and random). We end up with the following expression :

$$\begin{aligned}\log p((y, t_e, B_e), z, \theta \mid \Pi) = & \quad \log p(y \mid z, \theta, \Pi) \\ & + \log p(t_e, B_e \mid z, \theta, \Pi) \\ & + \log p(z_{re} \mid z_{fe}, \theta, \Pi) \\ & + \log p(z_{fe} \mid \theta, \Pi)\end{aligned}$$

The different log-likelihood parts are described below associated with their different assumptions. Note that the total formula of the log-likelihood is available in appendix A.1.

Longitudinal data attachment For longitudinal process modelling, we assumed that patients and their visits were independent conditionally to the parameters of the model and that the noise of the process follows a Gaussian distribution. We thus get:

$$\log p(y \mid z, \theta, \Pi) = \sum_{i,j} \log p(y_{i,j} \mid z, \theta, \Pi) = \sum_{i,j} -\log\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2\sigma^2}\left(y_{i,j} - \gamma_i\left(t_{i,j}\right)\right)^2$$

Survival data attachment For survival process modelling, we also assumed that all patients were independent and that the modelling of the survival process depended on whether the event was observed or not:

$$\begin{aligned}\log p(t_e, B_e \mid z, \theta, \Pi) = & \quad \textstyle\sum_i \log p(t_{e_i}, B_{e_i} \mid z, \theta, \Pi) \\ = & \quad \textstyle\sum_i \mathbb{1}_{B_{e_i}=1} \times \log\left(h_i(t_{e_i})\right) + \sum_i \log\left(S_i(t_{e_i})\right)\end{aligned}$$

Note that if $\psi_i(t) < t_0$ $\log\left(h_i(t_{e_i})\right) = -\infty$. To prevent estimation issues, we initialised the algorithm at a possible point getting inspiration from barrier methods (Nesterov, 2018).

Latent random effects priors attachment As patients were supposed to be independent of each other, we supposed that random effects were independent conditionally to z_{fe}, θ and Π. The regularization term associated, with $\bar{\xi} = 0$ and $\bar{\tau} = t_0$, is then:

$$\begin{aligned}\log p(z_{re} \mid z_{fe}, \theta, \Pi) = & \quad \log p(\tau_i \mid z_{fe}, \theta, \Pi) + \log p(\xi_i \mid z_{fe}, \theta, \Pi) \\ = & \quad -N \log\left(\sigma_\tau\sqrt{2\pi}\right) - \tfrac{1}{2\sigma_\tau^2}\textstyle\sum_i(\tau_i - t_0)^2 \\ & -N \log\left(\sigma_\xi\sqrt{2\pi}\right) - \tfrac{1}{2\sigma_\xi^2}\textstyle\sum_i(\xi_i - \bar{\xi})^2\end{aligned}$$

Latent fixed effects prior attachment Each latent fixed effect is independently sampled from a posterior distribution. The regularization term associated is then:

$$\begin{aligned}
\log p(z_{fe} \mid \theta, \Pi) = &\quad \log p(\tilde{g} \mid \theta, \Pi) + \log p(\tilde{v_0} \mid \theta, \Pi) \\
+ &\quad \log p(\tilde{\nu} \mid \theta, \Pi) + \log p(\tilde{\rho} \mid \theta, \Pi) \\
= &\quad -\log\left(\sigma_{\tilde{g}}\sqrt{2\pi}\right) - \tfrac{1}{2\sigma_{\tilde{g}}^2}\left(\tilde{g} - \overline{\tilde{g}}\right)^2 \\
- &\quad \log\left(\sigma_{\tilde{v}_0}\sqrt{2\pi}\right) - \tfrac{1}{2\sigma_{\tilde{v}_0}^2}\left(\tilde{v}_0 - \overline{\tilde{v}_0}\right)^2 \\
- &\quad \log\left(\sigma_{\tilde{\nu}}\sqrt{2\pi}\right) - \tfrac{1}{2\sigma_{\tilde{\nu}}^2}\left(\tilde{\nu} - \overline{\tilde{\nu}}\right)^2 \\
- &\quad \log\left(\sigma_{\tilde{\rho}}\sqrt{2\pi}\right) - \tfrac{1}{2\sigma_{\tilde{\rho}}^2}\left(\tilde{\rho} - \overline{\tilde{\rho}}\right)^2
\end{aligned}$$

3.2.4 Estimation of the Joint Temporal model

The estimation method can be separated into two steps, as in section 2.2.5.

The first one is the estimation of the fixed effects jointly with the random effects, referred to as the calibration step. It is the same as the one used in section 2.2.5. The adapted total log-likelihood, the sufficient statistics, and the maximisation update rules, necessary for the computation, are given in the appendix A.2 and A.3.

The second one is the estimation of the random effects of a patient knowing the fixed effects, named personalisation step. The only difference with section 2.2.5 is that for predictions, the survival probability is then corrected using the survival probability at the last visit as in other packages (Rizopoulos, 2016).

An implementation of the Joint temporal model is available in the open-source package Leaspy (v2): `https://gitlab.com/icm-institute/aramislab/leaspy`.

3.2.5 Attenuation estimation bias

An attenuation in the population speed v_0 has been observed in several studies using the Spatiotemporal model through Leaspy package (unpublished data). The cause of the attenuation of the speed v_0 is still a subject of research, although many hypotheses have been tested. As it is linked with the speed, ξ_i might also be linked to the issue. Thus, this attenuation issue might propagate to the scale of the Weibull distribution, ν, in the Joint Temporal model. Nevertheless, as $e^{-\xi_i}$ impacts ν this might result in an overestimation of ν. Here, we proposed a method to correct this attenuation bias for both the Longitudinal and the Joint Temporal model.

Past experiments have shown that predictions were less impacted by this attenuation. Thus, we decided to correct the fixed effects estimated by the calibration with the mean of the random effects estimated after a personalisation on the training set. We further investigate the size of the bias and the impact of the correction method.

3.2.6 Reference models for univariate longitudinal data and event with right non informative censoring

We chose to benchmark the Joint Temporal model against several reference models. First, we used one-process-only models. For the survival model, we used a Weibull AFT model

to describe the survival process, using the Lifelines python package (Davidson-Pilon, 2023). This model will be referred to as the AFT model. For the longitudinal model, we used a univariate version of the Spatiotemporal model described in section 3.2.2.2 using the open-source Leaspy package. This model will be referred to as the Longitudinal model. We expected the Joint Temporal model to be at least as good as these two models.

Second, we built a two-stage model: a survival model that took as covariates the random effects of the Longitudinal model (Murawska, Rizopoulos, and Lesaffre, 2012). Even though this model is subject to an immortal bias, it enabled us to compare our model to a better survival model than the AFT model. We used the Longitudinal model to extract random effects for each individual, and then used them as covariates in a Weibull AFT model, using the Lifelines package (Davidson-Pilon, 2023). This model will be referred to as the Two-stage model, and the Joint Temporal model was expected to be at least as good as it.

Third, we used a joint model with shared random effects, to evaluate if the new model could improve estimation. To do so, we use a logistic longitudinal process, estimated using the JMbayes2 package (Rizopoulos, 2016). This model will be referred to as the univ-JMbayes2 model. All the parameters and the model structure equations are summarised in Table 3.1.

3.2.7 Clinically meaningful estimated parameters

Predictions of different models can easily be compared. Nevertheless, joint models are also used to make cohort descriptions. Since each model has its own parameters, their comparison is difficult. To overcome this issue, the idea was to extract clinically meaningful estimated parameters that could be easily used in cohort description. For the survival process, we chose the median survival time, referred to as the median. For the longitudinal process, we have assumed a logistic shape of the curve. We then used the axis value and the slope at the inflection point as clinically meaningful estimated parameters. We later refer to them as midpoint and growth. Note that for all the different models, we reported the formula used in Table 3.2.

TABLE 3.1: Specification of the Joint Temporal model and reference models for univariate longitudinal outcome and event with non-informative right censoring

Legend: Longitudinal: Spatiotemporal longitudinal model, Two-stage model: AFT survival model that used as covariate random effects of the Longitudinal model, AFT: Accelerated Failure Time model, Joint Temporal: the Joint Temporal model, univ-JMbayes2: joint model with shared random effects. $spl(t)$: spline function, $x = (\xi, \tau)$ are extracted from the Longitudinal model and used as covariates,

Model	Inputs	Effects		Random effects structure	Link functions	
		Fix	Random	$\psi_i(t)$	Longitudinal $\gamma_i(t)$	Survival $S_i(t)$
Longitudinal	t	g, v_0, t_0	ξ_i, τ_i	$e^{\xi_i}(t-\tau_i)+t_0$	$\left(1+g\times\exp(-v_0\frac{(g+1)^2}{g}(\psi_i(t)-t_0))\right)^{-1}$	-
Two-stages	$t, x=(\xi_i,\tau_i)$	ν_0, ν_1, ρ_0	-	-	-	$exp\left(-\left(\frac{t}{\exp(\nu_0+\nu_1 x)}\right)^{\exp(\rho_0)}\right)$
AFT	t	ν_0, ρ_0	-	-	-	$exp\left(-\left(\frac{t}{\exp(\nu_0)}\right)^{\exp(\rho_0)}\right)$
Joint Temporal	t	g, v_0, t_0, ρ, ν	ξ_i, τ_i	$e^{\xi}(t-\tau)+t_0$	$\left(1+g\times\exp(-v_0\frac{(g+1)^2}{g}(\psi_i(t)-t_0))\right)^{-1}$	$exp\left(-\left(\frac{\psi_i(t)}{\nu}\right)^{\rho}\right)$
univ-JMBayes2	t	$\beta_0, \beta_1, g, \alpha, spl(t)$	$b_{i,0}, b_{i,1}$	$(\beta_0+b_{i,0})+(\beta_1+b_{i,1})t$	$(1+g\times\exp(\psi_i(t)))^{-1}$	$\exp\left(-\int_0^t \exp(spl(u)+\alpha\gamma(u))du\right)$

TABLE 3.2: Clinically meaningful estimated parameters formulas for the Joint Temporal model and reference models for univariate longitudinal outcome and event with non-informative right censoring

Legend: Longitudinal: Spatiotemporal longitudinal model, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, AFT: Accelerated Failure Time model, Joint Temporal: the Joint Temporal model, univ-JMbayes2: joint model with shared random effects. If a function to extract the clinically meaningful estimated parameters was available in the package it was directly used and noticed as available in the package. For the JMBayes2 package, we estimated the function S(t) to find the median of survival (from S(t) estimation).

Model	Survival	Longitudinal	
	Median	Growth	Midpoint
Longitudinal	-	$v_0\frac{(g+1)^2}{4g}$	$\log(g)\frac{g}{v_0(g+1)^2}+t_0$
Two-stages	available in package	-	-
AFT	available in package	-	-
Joint Temporal	$\nu\ln(2)^{\frac{1}{\rho}}$	$v_0\frac{(g+1)^2}{4g}$	$\log(g)\frac{g}{v_0(g+1)^2}+t_0$
univ-JMBayes2	from S(t) estimation	$-\beta_1/4$	$-\beta_0$

3.3 Simulation study

3.3.1 Simulated data

3.3.1.1 Data simulation

Data were simulated under the Joint Temporal model structure. We used the following procedure:

1. We simulated random effects using $\xi_i \sim \mathcal{N}\left(0, \sigma_\xi^2\right)$ and $\tau_i \sim \mathcal{N}\left(t_0, \sigma_\tau^2\right)$.
2. We modelled the age at baseline $t_{b,i}$ as $t_{b,i} = \tau_i + \delta_{b_i}$ with $\delta_{b_i} \sim \mathcal{N}\left(\overline{\delta_b}, \sigma_{\delta_b}^2\right)$.
3. We set a time of follow-up per patient T_{f_i}, with $T_{f_i} \sim \mathcal{N}\left(\overline{T_f}, \sigma_{T_f}^2\right)$ and a time between two visits $\delta_{v_{i,j}} = t_{i,j-1} - t_{i,j}$, with $\delta_{v_{i,j}} \sim \mathcal{N}\left(\overline{\delta_v}, \sigma_{\delta_v}^2\right)$ to simulate n_i visits until $t_{i,n_i} \leq t_{b,i} + T_{f_i} < t_{i,n_{i+1}}$.
4. We set the value of the outcome at each visit using $y_{i,j} = \gamma_i(t_{i,j}) + \epsilon_{i,j}$ with $\epsilon_{i,j} \sim \mathcal{N}\left(0, \sigma^2\right)$ with Leaspy.
5. For each patient, we simulated the event T_{e_i} using $T_{e_i} \sim e^{-\xi_i}\mathcal{W}\left(\nu, \rho\right) + \tau_i$.
6. We considered that the event stopped the follow-up and that the follow-up censored the event. Thus all the visits after the event were censored ($t_{i,j} > T_{e_i}$) and events after the last visit were censored: ($t_{i,max(j)} < T_{e_i}$).

Due to random assignment, some patients experienced the event (death) before their first visit. This censure also changed the random effects distribution, as most of the patients censored were progressing fast. To consider this bias, we corrected the true values of clinically meaningful estimated parameters, using the available true value of random effects.

3.3.1.2 Simulated Scenarios

We simulated ALS real-like data with different data collection designs (number of patients, visit frequency, time of follow-up, and noise). We used the PRO-ACT dataset, described in section 2.3.1, to get estimated real-like values for parameters. Parameters directly associated with the disease were extracted from data analysis, using the Longitudinal and AFT models. For each of the five scenarios that studies the impact of a given parameter, we simulated three different datasets, that corresponded to an easy, medium, and complex situation.

Our first scenario concerned the number of patients needed for calibration. Due to PRO-ACT properties, we decided to test three sample sizes: N = 200, 500, and 1,000, with other parameters being equal to real-like ones. Visit frequency impact was studied with our second scenario. In PRO-ACT, patients are seen on average every 1.47 months. We have tested three different average times between visits $\overline{\delta_v}$ = 3, 1.5, 1. The third scenario enabled us to study the impact of the time of follow-up. Patients in PRO-ACT are followed for an average of 0.96 years. We tested three different average follow-up duration $\overline{T_f}$ = 0.75, 1.0, 1.25. Our fourth scenario enabled us to measure the impact of longitudinal noise in data collection. On PRO-ACT, we found a noise of 5% on the longitudinal outcome using the

Longitudinal model. Note that the test-retest error of ALSFRSr is around 2% (Miano et al., 2004). We thus decided to test $\sigma = 5\%$, 10% and 20%. Finally, our fifth scenario concerned noise in event distribution. We found a shape of the Weibull distribution, ρ, on the survival process of 2.25. We decided to test three events distribution $(\rho, \nu) = (3, 3.11), (2.25, 3.62), (2, 3.84)$ so that the mode of the distribution $(\nu \left(\frac{\rho-1}{\rho}\right)^{\frac{1}{\rho}})$ does not change but the variation increases.

To make the results of our scenarios comparable, we put the same real-like dataset for all the medium difficulties of the scenarios except for the patient and longitudinal noise one, for which the real-like dataset corresponds to the easier one. We ended up with 11 distinct datasets for different study designs, all inspired by real data for which simulation parameters are summarised in the Table 3.3.

TABLE 3.3: Data simulation parameters for each of the five scenarios

Legend: Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. (r) indicate when ALS real-like parameters are used. For each scenario, three datasets were simulated [Easy, Medium, Hard] that corresponded to an increased difficulty of simulated data for the estimation. More information on simulation is available in section 3.3.1.2. Note that to assure identifiability: $\overline{\tau} = t_0, \overline{\xi} = 0$.

Type	Parameters Name		Symbol	ALS real-like (r)	Patient number (P)	Visit density(V)	Follow up (F)	Longitudinal noise (L)	Survival noise (S)
Patients	Patient number		N	1,000	[r, 500, 200]	r	r	r	r
Random Effect	Individual log-speed factor	(mean)	$\overline{\xi}$	0	r	r	r	r	r
		(std)	σ_ξ	0.73	r	r	r	r	r
	Population estimated reference time (years)	(std)	$\overline{\tau}$	1.17	r	r	r	r	r
		(std)	σ_τ	1.04	r	r	r	r	r
Longitudinal Fixed Effects	Speed of the logistic curve		v_0	1.13	r	r	r	r	r
	Curve value at t_0: $\frac{1}{1+g}$		g	6.40	r	r	r	r	r
	Estimated noise		σ	0.04	r	r	r	[r, 0.1, 0.2]	r
Survival Fixed Effects	Scale of the Weibull distribution		ν	3.62	r	r	r	r	[3.11, r, 3.84]
	Shape of the Weibull distribution		ρ	2.25	r	r	r	r	[3, r, 2]
Visits	Time between τ and baseline (years)	(mean)	$\overline{\delta_b}$	0.4	r	r	r	r	r
		(std)	σ_{δ_b}	0.84	r	r	r	r	r
	Time of follow up (years)	(mean)	$\overline{T_f}$	0.96	r	r	[2, r, 0.5]	r	r
		(std)	σ_{T_f}	0.87	r	r	r	r	r
	Time between visits (months)	(mean)	$\overline{\delta_v}$	1.47	r	[1, r, 3]	r	r	r
		(std)	σ_{δ_v}	0.5	r	r	r	r	r

3.3.2 Method - Validation study

3.3.2.1 Sensitivity analysis

To perform our sensitivity analysis on all the simulated scenarios (Table 3.3 in appendix), we first trained the Longitudinal model during 2,000 iterations. We used the values of the Longitudinal model as initialisation for the Joint Temporal model, and ran it for 70,000 iterations (on average an hour), with a Robbins-Monro convergence phase on the 10,000 last iterations (Robbins and Monro, 1951). We then assessed the quality of the estimates in two steps.

Effects estimations We wanted to assess how well the model could estimate fixed and random effects from simulated data whatever the study design. For fixed effects, as done by Pan and Hout, 2023, we extracted data from the posterior distribution (between 30,000 and 60,000 iterations) and estimated the bias to true fixed effects used for simulation. For random effects, we extracted the mean of the random effects posteriors for each individual (between 30,000 and 60,000 iterations) to reduce the computation complexity. As a metric, we used the intraclass correlation between the mean of each individual and the true value that enabled the simulation.

Attenuation bias correction We tested our correction of attenuation bias in simulated data, using clinically meaningful estimated parameters, as their biases are easier to interpret than fixed effects. For the value without correction, we computed the clinically meaningful parameters using the posteriors of fixed effects. For the value with correction, we corrected the output on the model ($70,000^{th}$ iteration, means computed through the Robbins-Monro convergence), with the mean of random effects estimated thanks to personalisation on the training set. We compared both values.

The impact of the correction method on random effects was also evaluated. We computed the intraclass correlation between the random effects extracted from personalisation and the true value that enabled the simulation. Results were compared with the one of the sensitivity analysis.

3.3.2.2 Benchmark experiment on simulated data

All estimations and predictions of the models were compared using the real-like dataset (Table 3.3 in appendix). We selected patients with at least 3 visits to ease the following prediction setup. A 10-fold cross-validation (train 90% - test 10%) was used, with 70,000 iterations for both the Spatiotemporal and Joint Temporal model (on average half an hour) with a Robbins-Monro convergence phase on the 10,000 last iterations (Robbins and Monro, 1951). The Joint Temporal model initialised with parameters of a Longitudinal model ran with 2,000 iterations.

Clinically meaningful estimated parameters on simulated data As we could not compare the fixed effects of the different models, we chose to compare the clinically meaningful estimated parameters, with the true one that enabled simulation.

For each model, we computed clinically meaningful estimated parameters using model output ($70,000^{th}$ iteration, means computed through the Robbins-Monro convergence) (Table 3.2) and applied the attenuation bias correction (see section 3.2.5) for both the Joint Temporal model and the Longitudinal model. Then, the bias to true clinically meaningful estimated parameters was computed. We compared the 10-fold pooled bias distribution of the Joint Temporal model with the four reference models for the three clinically meaningful estimated parameters using a Wilcoxon signed-rank test with a Bonferroni adjustment for multiple pairwise comparisons.

Prediction on simulated data Finally, we wanted to evaluate both the goodness of survival and longitudinal predictions of the Joint Temporal model against reference models on simulated data. The fixed effects output by the model ($70,000^{th}$ iteration) were used as a predictive model and will be referred to as the prediction model. Personalisation was made using the first two visits of the patients from the test set and predictions were estimated on the remaining. The goodness of longitudinal predictions was assessed using absolute errors. The goodness of survival predictions was assessed using the C-index, for event order at 1 and 1.5 years, as in the ALS challenge (Kueffner et al., 2019). Nevertheless, this method is not suitable for the evaluation of t-year predicted risks (Blanche, Kattan, and Gerds, 2019). We thus added the mean cumulative dynamic AUC at 1 and 1.5 years for a correct measure and kept the C - index for comparison with existing results. To evaluate the absolute distance between real and estimated probability of the events, we used the Integrated Brier Score (IBS). The predictions were compared using a Wilcoxon signed-rank test with a Bonferroni adjustment for multiple pairwise comparisons. All the survival metrics were computed using the Python package sksurv (Pölsterl, 2020).

3.3.3 Results

3.3.3.1 Data simulated

Among the 11 datasets simulated under the 5 scenarios, the number of visits simulated was relatively stable at 8,341 (31) except for the patient, visits and follow-up scenarios (Figure B.1 in appendix). The censuring rate was between 80 and 85% for all datasets, except for the easy and hard dataset of the follow-up scenario, with 60% and 90% censure (Figure B.1 in appendix).

3.3.3.2 Sensitivity analysis

Effect estimation We evaluated the posterior distribution of fixed effects (Figure B.2 in appendix). For the survival submodel, both Weibull fixed effects were slightly overestimated. The scale of the Weibull distribution (ν) was overestimated for the real-like dataset by 8.7 (2.9) % and reached the worst results for the harder dataset of the follow-up scenario (Figure B.2 in appendix). The shape of the Weibull distribution (ρ) was overestimated for the real-like dataset by 30.2 (5.5) % and often across scenarios (Figure B.2 in appendix).

For the longitudinal submodel, the main parameters were slightly underestimated. For the real-like dataset, g was underestimated by -5.2 (3.9) % (with $\frac{1}{1+g}$ the value of the

outcomes at t_0) and the population speed (v_0) by -6.7 (2.8)%. The worst results were reached for the harder dataset of longitudinal noise scenarios (Figure B.2 in appendix).

For the four remaining parameters, the population estimated reference time (t_0), the longitudinal noise (σ) and the standard deviation of the random effects (σ_τ, σ_ξ), average biases were below 4% for the real-like dataset. Worst results were reached for the harder dataset of longitudinal noise scenarios (σ = 20%) for all these fixed effects (Figure B.2 in appendix).

The intra-class correlation between simulated random effects and the mean of posterior random effects for each patient was between 0.75 and 0.97 (Table B.1 in appendix).

Attenuation bias correction The attenuation correction bias was evaluated on clinically meaningful estimated parameters, but was already observed in effect estimation with the underestimation of the population speed (v_0) and the overestimation of the scale of the Weibull distribution (ν). Almost all corrected clinically meaningful estimated parameters had a lower bias compared to those of the posterior, whatever the scenario and difficulty of the dataset (Figure 3.1).

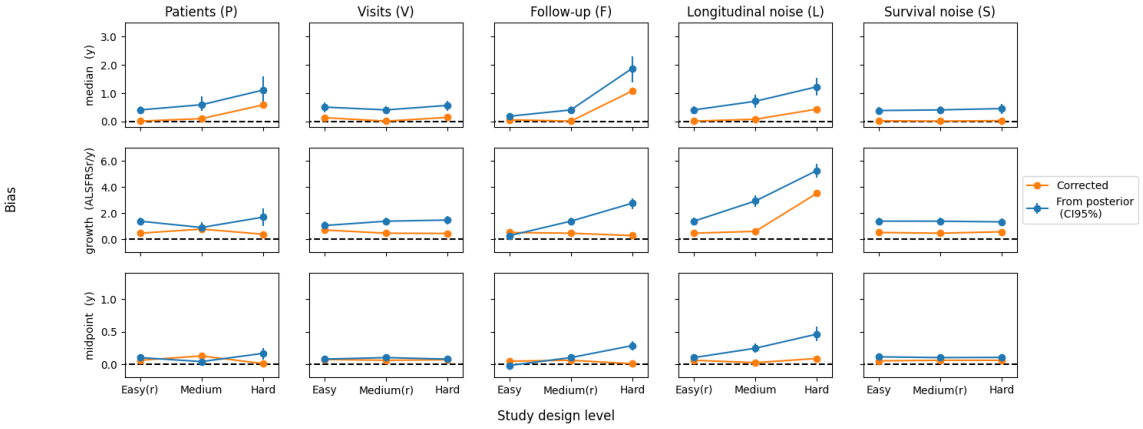


FIGURE 3.1: Absolute estimated bias for clinically meaningful estimated parameters for the Joint Temporal model (median, growth and midpoint) on all simulated datasets

Legend: Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. Easy, Medium, and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in section 3.3.1.2 and summarised in the table 3.3. Datasets with (r) correspond to the real-like dataset.
corrected: mean bias extracted from Robbin Monroe algorithm corrected for attenuation bias, from posterior: mean bias with CI 95% of the posterior chain. Values of the clinically meaningful estimated parameters that enabled the simulation, from which the bias is computed, were 4.48 (0.12) years for the median, -13.19 (0.21) point of ALSFRSr per year for the growth, 2.92 (0.06) years for the midpoint.

The corrected median absolute bias was below 0.43 years for a real value of 4.48 (0.12) years, for all scenarios except the harder of patients and follow-up scenarios. The worst result was reached for the harder follow-up scenario ($\overline{T_f}$ = 0.75). The corrected growth absolute bias was below 0.76 points for a real value of -13.19 (0.21) points of ALSFRSr per

year, for all scenarios, except for the harder longitudinal noise dataset (σ = 20%). The corrected absolute bias midpoint was lower than 0.13 years for a real value of 2.92 (0.06) years, whatever the scenario. The worst result was reached for the harder dataset of the longitudinal noise scenario (σ = 20%). All the detailed values for biases are available in Table B.2 in the appendix.

Random effects computed through personalisation were not different from those of the posterior mean (Figure B.3 in appendix). Detailed values can be found in Table B.3 in the appendix.

3.3.3.3 Benchmark experiment on simulated data

Clinically meaningful estimated parameters For the longitudinal process, the Joint Temporal and Longitudinal models were closer to the Truth model compared to univ-JMbayes2 (Figure B.4 A). Over 10-fold cross-validation, the midpoint was simulated to be 2.97 (0.02) years and the growth of -12.97 (0.11) points of ALSFRSr per year. The Joint Temporal model had a smaller absolute bias on both the growth and midpoint compared to the univ-JMbayes2 model: 0.59 (0.08) compared to 1.71 (0.12) points of ALSFRSr per year (p-value = 3.9e-03) and 0.06 (0.01) compared to 0.43 (0.04) years (p-value = 3.9e-03)(Table 3.4). This was expected due to the structure of the simulated data. The Longitudinal model had a smaller absolute bias on both the growth and the midpoint compared to the Joint Temporal model: 0.39 (0.08) compared to 0.59 (0.08) (p-value =3.9e-03) and 0.04 (0.01) compared to 0.06 (0.01) (Table 3.4). This might be because the Longitudinal model has fewer constraints and is applied to an easy task: the simulated data and the Longitudinal model have the same structure.

For the survival process, the Two-stages model and the Joint Temporal model were close to the Truth at the beginning of the progression. The univ-JMbayes2 model overestimated the average patient survival (Figure B.4 B in appendix). It was confirmed by an absolute bias of 2.34 (0.14) years on the median compared to a truth value of 4.57 (0.03) years over 10-fold cross-validation (Table 3.4). This issue might be linked to the non-linearity of the random effects in the simulation process. The computed bias on the median was not significantly different for the Joint Temporal model and the Two-stages model. The computed bias on the median was lower for the Joint Temporal model than for the AFT model: respectively 0.09 (0.04) and 0.8 (0.08) (p-value = 5.9e-03) (Table 3.4).

Next visits predictions on simulated data On simulated data, using the 10-fold cross-validation (train 90% - test 10%), we personalised the model on the first two visits of the patients from the test set and predicted the remaining. Doing so, 6,382 longitudinal predictions were made at 0.58 (0.41) years from the last visit.

For the longitudinal process, the Joint Temporal model had a significantly smaller longitudinal absolute prediction bias compared to the univ-JMbayes2 model: respectively 3.74 (3.8) and 4.09 (4.1) points of ALSFRSr (p-value = 3.9e-27)(Table 3.5). The Longitudinal model had a smaller absolute prediction bias compared to the Joint Temporal model: respectively 3.73 (3.82) compared to 3.74 (3.8) (p-value = 3.3e-05).

TABLE 3.4: Clinically meaningful estimated parameters absolute bias of the Joint Temporal model and reference models on real-like dataset

Legend: Joint Temporal: the Joint Temporal model, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, Longitudinal: Spatiotemporal longitudinal model, AFT: Accelerated Failure Time model, univ-JMbayes2: joint model with shared random effects.
Results are presented with the mean (SD) over the 10-fold cross-validation. P-values are computed using a Wilcoxon signed-rank test with Bonferroni correction between the Joint Temporal model and each reference model. Results in bold are the smallest biases for the clinically meaningful estimated parameter. Values of the clinically meaningful estimated parameters that enabled the simulation, from which the biases were computed: -12.97 (0.11) points of ALSFRSr per year for the growth, 2.97 (0.02) years for the midpoint, 4.57 (0.03) years for the median.

	Joint Temporal	Two stages	p-value	Longitudinal	p-value	AFT	p-value	univ-JMbayes2	p-value
growth (pt/y)	0.59 (0.08)	-	-	**0.39 (0.08)**	3.9e-03	-	-	1.71 (0.12)	3.9e-03
midpoint (y)	0.06 (0.01)	-	-	**0.04 (0.01)**	3.9e-03	-	-	0.43 (0.04)	3.9e-03
median (y)	**0.09 (0.04)**	0.14 (0.05)	3.2e-01	-	-	0.80 (0.08)	5.9e-03	2.34 (0.14)	5.9e-03

For the survival process, the Joint Temporal model significantly improved all metrics compared to the univ-JMbayes2 and AFT models. Again, for the univ-JMbayes2 model, this might be due to the failure of the univ-JMbayes2 model to handle the non-linear relation between random effects. The survival predictions of the Joint Temporal model were not significantly distinct for any metric from the one of the Two-stages model.

TABLE 3.5: Prediction metrics results for the Joint Temporal model and reference models on real-like dataset

Legend: Joint Temporal: the Joint Temporal model, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, Longitudinal: longitudinal model, AFT: Accelerated Failure Time model, univ-JMbayes2: joint model with shared random effects.
Results are presented with means (SD) over the 10-fold cross-validation. P-values are computed using a Wilcoxon signed-rank test or pairwise t-test with Bonferroni between the Joint Temporal model and each reference model. IBS stands for Integrated Brier Score. ↓means that the metric should be minimised and ↑maximised. Results in bold are the best for each metric. 6,382 longitudinal predictions were made at 0.58 (0.41) years from the last visit.

	Joint Temporal	Two stages	p-value	Longitudinal	p-value	AFT	p-value	univ-JMbayes2	p-value
Absolute bias ↓	3.74 (3.80)	-	-	**3.73 (3.82)**	3.3e-05	-	-	4.09 (4.10)	3.9e-27
MSE	28.42 (70.5)	-	-	28.54 (72.36)	4.4e-02	-	-	33.48 (83.35)	3.0e-25
IBS ↓	**0.08 (0.02)**	0.08 (0.02)	3.2e-01	-	-	0.11 (0.03)	5.3e-05	0.11 (0.03)	7.5e-05
Mean AUC (1y, 1.5y) ↑	0.74 (0.10)	**0.75 (0.09)**	1.0e+00	-	-	0.60 (0.08)	2.2e-03	0.60 (0.09)	3.3e-03
C-index 1.0y ↑	**0.73 (0.07)**	0.73 (0.08)	1.0e+00	-	-	0.60 (0.07)	3.1e-03	0.59 (0.07)	2.5e-03
C-index 1.5y ↑	**0.73 (0.07)**	0.73 (0.07)	1.0e+00	-	-	0.60 (0.07)	2.0e-03	0.59 (0.07)	1.9e-03

3.3.4 Intermediate conclusion

We have evaluated the impact of attenuation bias on the Joint temporal model fixed effects and proposed a correction that enabled us to reduce the absolute bias on clinically meaningful estimated parameters. On simulated data, the Joint Temporal model got results close to the Longitudinal model. Results in the survival process were in the same range as the ones of the Two-stage model and outperformed the AFT model. All these results were promising and enabled us to validate the Joint Temporal model on simulated data.

3.4 Application: Model benchmark

3.4.1 ALS Data: PRO-ACT

Study population We applied our method to the PRO-ACT data described in section 2.3.1 to test prediction performances on real data keeping in mind the objective to improve the power of clinical trials. For the study, we selected patients with at least three visits for the longitudinal outcomes described below (to ease prediction setup), age at first symptoms, symptom onset (spinal or bulbar), and sex.

Outcomes As described by the FDA (FDA, 2019), for ALS, clinical trials must demonstrate a treatment effect on function in daily activities and death. Following this guideline, we considered the revised version of the ALS functional rating scale revised (ALSFRSr) as the longitudinal outcome (Rooney et al., 2017). The scale starts at a maximum theoretical value of 48 and decreases with the severity of the disease till zero. For modelling reasons, it was normalized using its scale so that the 0 value was the healthiest score and +1 the maximum disease score change. Tracheostomy was associated with death as a survival outcome, as in many ALS studies and challenges (Kueffner et al., 2019).

3.4.2 Method - Benchmark experiment on real data

We compared the estimation and prediction of the reference models on real data, using a 10-fold cross-validation (train 90% - test 10%). The Joint Temporal model was run with 70,000 iterations for each model (on average an hour and a half) with a Robbins-Monro convergence phase on the 10,000 last iterations (Robbins and Monro, 1951). The Joint Temporal model was initialised with parameters of a Longitudinal model ran for 2,000 iterations.

Clinically meaningful estimated parameters on real data We computed corrected clinically meaningful estimated parameters from the Joint Temporal model and the reference models to see how different they could be (see section 3.3.2.1).

Prediction on real data We wanted to evaluate both the goodness of survival and longitudinal predictions of the Joint Temporal model against reference models, this time on real data using the PRO-ACT database. We used the same method as the one developed in section 3.3.2.2, but applied it to PRO-ACT data: a 10-fold cross-validation (train 90% - test 10%), with personalisation of the two first visits of the patients of the tests set to predict the remaining.

3.4.3 Results

3.4.3.1 Data description

PRO-ACT datasets had similar or easier characteristics for estimation, compared to our real-like simulated dataset, in terms of the number of patients, visits, and time of follow-up (Table 3.6).

TABLE 3.6: Characteristics of the selected patients of PRO-ACT database used to benchmark the Joint Temporal model with the reference models

Legend: Results are presented as mean (SD) or frequency [percent][class%]. There were no missing values in the dataset due to patient selection.

Characteristics	Values
Number of patients	2,528
Number of visits	23,143
Number of patients years	2,531
Percentage of censored events (%)	76.74
Number of visits per patients	9.2 (4.3)
Time of follow-up (years)	1.0 (0.6)
Time between visits (months)	1.5 (0.9)
Gender (Male)	1,575 [62.3%]
Symptom onset (Spinal)	1,952 [77.2%]
Age at first symptoms	54.0 (11.3)
Time from first symptoms to baseline (years)	1.6 (0.9)
ALSFRSr total (baseline)	37.9 (5.4)

3.4.3.2 Benchmark on real data

Clinically meaningful estimated parameters on real data For the longitudinal process, the Joint Temporal model and Longitudinal model were close compared to the univ-JMbayes2 model (Figure B.5 A in appendix). Growth in points of ALSFRSr per year ranged from -13.95 (0.08) for the Longitudinal model to -15.351 (0.1) for the univ-JMbayes2 model, with a difference of almost 1.4 points per year between the two models. The midpoint ranged from 2.78 (0.01) for the Joint Temporal model to 3.22 (0.014) for the univ-JMbayes2 model, with a difference of more than 5 months (Table 3.7).

Survival curves were pretty close, even though the variation over 10-fold cross-validation was wider for the univ-JMbayes2 model (Figure B.5 B in appendix). Apart from the AFT model, the computed medians were close for all the models, ranging from 4.14 (0.02) for the Two-stages model to 4.47 (0.09) for the univ-JMbayes2 model with a difference of almost 4 months between the two models (Table 3.7).

TABLE 3.7: Clinically meaningful estimated parameters for the Joint Temporal model and reference models on PRO-ACT data

Legend: Joint Temporal: the Joint Temporal model, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, Longitudinal: Spatiotemporal longitudinal model, AFT: Accelerated Failure Time model, univ-JMbayes2: joint model with shared random effects.
Results are presented with the mean (SD) over the 10-fold cross-validation. P-values are computed using a Wilcoxon signed-rank test with Bonferroni correction between the Joint Temporal model and each of the reference models.

	Joint Temporal	Two stages	p-value	Longitudinal	p-value	AFT	p-value	univ-JMbayes2	p-value
growth (pt/y)	-14.17 (0.08)	-	-	-13.95 (0.08)	9.8e-11	-	-	-15.35 (0.10)	4.6e-14
midpoint (y)	2.78 (0.01)	-	-	2.79 (0.01)	1.6e-04	-	-	3.22 (0.01)	1.8e-13
median (y)	4.15 (0.03)	4.14 (0.02)	6.2e-01	-	-	4.69 (0.02)	4.4e-11	4.47 (0.09)	2.0e-06

Prediction on real data On real data, using the 10-fold cross-validation (train 90% - test 10%), we personalised the model on the first two visits of the patients from the test set and predicted the remaining. Doing so,18,077 longitudinal predictions were made at 0.63 (0.55) years from the last visit.

For the longitudinal process, the Joint Temporal model had a significantly smaller absolute prediction bias compared to univ-JMbayes2 respectively 4.21 (4.41) and 4.24 (4.14) points of ALSFRSr (p-value = 1.4e-17), but a larger absolute prediction bias compared to the Longitudinal model (4.18 (4.38) (p-value = 2.2e-73)) (Table 3.8).

For the survival process, the Joint Temporal model was significantly better than all the other models for ordering events, with a mean AUC of 0.67 (0.07) (Table 3.8). The distance to the observed failure time was not significantly different from the one of the univ-JMbayes2 model with an IBS of 0.1 (0.01), but it was significantly smaller than the one of the Two-stage model (0.11 (0.01) (p-value = 2.5e-04)) and the one of the AFT model (0.12 (0.01) (p-value = 1.1e-04)) (Table 3.8).

TABLE 3.8: Prediction metrics results for the Joint Temporal model and reference models on PRO-ACT data

Legend: Joint Temporal: the Joint Temporal model, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, Longitudinal: Spatiotemporal longitudinal model, AFT: Accelerated Failure Time model, univ-JMbayes2: joint model with shared random effects.
Results are presented with the mean (SD) over the 10-fold cross-validation. P-values are computed using a Wilcoxon signed-rank test or pairwise t-test with Bonferroni between the Joint Temporal model and each of the reference models. IBS stand for Integrated Brier Score. ↓means that the metric should be minimised and ↑maximised. Results in bold are the best for each metric. 18,077 longitudinal predictions were made at 0.63 (0.55) years from the last visit.

	Joint Temporal	Two stages	p-value	Longitudinal	p-value	AFT	p-value	univ-JMbayes2	p-value
Absolute bias ↓	4.21 (4.41)	-	-	**4.18 (4.38)**	2.2e-73	-	-	4.24 (4.14)	1.4e-17
MSE ↓	37.11 (91.07)	-	-	36.67 (90.11)	1.4e-63	-	-	**35.14 (71.33)**	1.1e-13
IBS ↓	**0.10 (0.01)**	0.11 (0.01)	2.5e-04	-	-	0.12 (0.01)	1.1e-04	0.10 (0.01)	1.0e+00
Mean AUC (1y, 1.5y) ↑	**0.67 (0.07)**	0.62 (0.08)	2.8e-03	-	-	0.42 (0.07)	4.3e-05	0.61 (0.09)	1.7e-03
C-index 1.0y ↑	**0.69 (0.05)**	0.63 (0.06)	4.0e-04	-	-	0.41 (0.05)	1.1e-06	0.63 (0.06)	3.1e-04
C-index 1.5y ↑	**0.70 (0.05)**	0.65 (0.05)	9.7e-04	-	-	0.41 (0.05)	1.6e-06	0.66 (0.05)	1.8e-03

3.5 Discussion

3.5.1 Results interpretation

3.5.1.1 Methodology

Our work showed the potential of a Joint Temporal model that automatically realigns patients' trajectories in the context of progressive chronic diseases. After validating the Joint Temporal model on simulated data, we showed how it could improve the prediction of ALS progression. Compared to the AFT model and the Two-stage model, the Joint Temporal model outperformed significantly all the metrics, but the Longitudinal model performed slightly better than the Joint Temporal model. Event censoring may not bring much to the Longitudinal model in this case, however, joint modelling corrected shortcomings of the two-stage model, such as immortal bias. Compared to the joint univariate model with shared

random effects (univ-JMbayes2 model), the Joint Temporal model outperformed the longitudinal and event ordering metrics (C-index and AUC). This shows that our latent disease age might be better suited to capture the heterogeneity of the progression of degenerative diseases. Note that for the event distance metric (IBS) the joint univariate model with shared random effects model and the Joint Temporal model did not perform significantly differently. This might be because the survival function of the joint univariate model with shared random effects model, was more flexible, using splines instead of a Weibull function.

Prediction performances can be assessed in real datasets, using hidden visits. The Longitudinal model gave the best longitudinal predictions. The clinically meaningful estimated parameters were close between the Longitudinal and the Joint Temporal models. The difference between the joint models on real data was 1.18 points of ALSFRSr per year on the slope and more than 5 months on the midpoint. These differences were wider than the absolute bias observed in the real-like simulated data. For survival, the Joint Temporal model outperformed the other models on the real dataset. Compared to the joint univariate model with shared random effects model, we observed a difference of almost 4 months for the median survival. This difference was of the same magnitude as the median absolute bias observed on the real-like simulated data (Table 3.4).

We have tested the Joint Temporal model in different clinical trial scenarios to give future users insights into biases on clinically meaningful estimated parameters and guidelines as recommended (Hickey et al., 2016). Note that we studied ALS disease with its own pathological and clinical study characteristics and that guidelines should be adapted in other contexts. We showed that the Joint Temporal model tends to overestimate the survival median and underestimate absolute growth with scenarios increasing difficulty (fewer points, larger noise). Still, the following recommendations could help keep this bias reasonably small. The parameter that impacted the estimation precision the most, was the noise on longitudinal data. In clinical scores, it could be assessed with the Minimal Detectable Change (MDC) and can vary from one disease to another: 10 % for MDS-UPDRS III (Parkinson), (Steffen and Seney, 2008), 10 % MMSE (Alzheimer) (Feeney et al., 2016) and 3% ALSFRSr (ALS) (Fournier, James, and Glass, 2023). In our simulation study, a longitudinal noise of 20% was too large to get good results, but good results were reached with 10 % in our setting (corrected bias: [median: 0.07 year, growth: -0.59 point of ALSFRSr/year, midpoint: -0.02 year]). The number of patients and the time of follow-up should also be considered. For ALS, with a 4 to 5 years median of death from first symptoms, 6 months of follow-up induces too much bias on the estimated median which might be due to a high censoring rate in our case (90%). Results with one year of follow-up, that correspond to real-like studies (83% of censoring rate), were good (corrected bias: [median: 0.005 year, growth: 0.45 point of ALSFRSr/year, midpoint: -0.06 year]). Note that observation time in disease timing might also play a role here but we did not further investigate it. We showed that even with only 200 patients, the Joint Temporal model had reasonable biases on clinically meaningful estimated parameters (corrected bias: [median: 0.58 year, growth: 0.37 point of ALSFRSr/year, midpoint: 0.01 year]). Note that there might be a balance with the number

of visits (8 by default here). For 1,000 patients, we could reduce the number of visits to 4 per patient every 3 months again with a reasonable bias (corrected bias: [median: 0.14 year, growth: 0.43 point of ALSFRSr/year, midpoint: -0.07 year]). Note that the visits were well-distributed during the whole follow-up time.

3.5.1.2 Clinical

In ALS the question of prognosis is of great interest both for patient care management and for the increase of clinical trial power. A prognosis online tool is freely available online for neurologists: The ENCALS Survival model (Westeneng et al., 2018) using a Royston Parmar Survival model with a proportional impact of covariates (see section 1.3). Using the C-index, the prediction of our Joint Temporal model reached a C-index around 0.7 (0.5) using only ALSFRSr, whereas the ENCALS model reached in external validation cohort 0·78 (95% CI 0·77–0·80), using various covariates as sex, site of onset or Forced Vital Capacity. The performances of the Joint Temporal model can also be compared to the results of the ALS challenges Prize4Life (Kueffner et al., 2019). The C-index estimated corresponds to rank seventh, even though we used only ALSFRSr progression to help the prediction, the other models being much more flexible models with multiple covariates. Longitudinal prediction got an absolute bias of 4.18 (4.38) points of ALSFRSr which is correct for a scale design of 48 points but that could still be improved compared to the Minimum Detectable Change (MDC) of 1.59 points of ALSFRSr(Fournier, James, and Glass, 2023). These results are quite encouraging given the quantity of information used for prediction but also call for an extension of the model, to improve the quantity of data handled.

3.5.2 Strengths & Weaknesses

3.5.2.1 Methodology

The main strength of our model, compared to the state-of-the-art is that it alleviates the proportional hazard hypothesis while keeping an interpretable structure with the latent disease age. Compared to the joint univariate model with the shared random effects model, which models the effect of longitudinal outcomes on the event, the Joint Temporal model captures a latent shared process with the latent disease age, which provides a modelling alternative. The shared latent age encapsulates the individual temporal variability (speed and reference time) and enables realigning the progression of the patients on a common timeline.

More work could be done to alleviate the assumptions about the independence of longitudinal noises at each visit, which has already been addressed in some models (Proust-Lima, Philipps, and Liquet, 2017). The Joint Temporal model has also been shown to be subjected to a bias that attenuates the speed, and although we attempted to address it, we still need to consolidate the hypothesis and the estimation of a good correction for parameters.

3.5.2.2 Clinical

As in the precedent chapter 2, the PRO-ACT database is not representative of the real-life ALS patient population (over-representation of men, young subjects, patients with spinal

onset, and slow progressors). Thus, prediction performances must be tested on cohorts representative of the clinical routine and most of all with external databases.

3.6 Conclusion & perspectives

In conclusion, the Joint Temporal model with latent disease age enabled us to improve the performance of most prediction metrics compared to existing joint models and alleviate the need for a precise reference time.

From a clinical point of view, this model opens up perspectives to design predictive and personalized therapeutic strategies. Nevertheless, it could only model one longitudinal and one event with non-informative right censoring, whereas in ALS many outcomes are available as well as events that could be censored by death as the initiation of Non-Invasive Ventilation (NIV).

Chapter 4

A Joint cause-specific Spatiotemporal model

In this chapter, we extended the univariate Joint Temporal model, into a multivariate Joint cause-specific Spatiotemporal model. This extension enabled us to analyse the initiation of Non-Invasive Ventilation, potentially censured by death, jointly with longitudinal scores linked to functional decline. We used the longitudinal Spatiotemporal model presented in Chapter 2 associated with a cause-specific Weibull distribution from the latent disease age. Spatial random effects were incorporated to have a proportional effect on the hazard. First, we validated our model on simulated real-like data. Then, we benchmarked our model with a reference joint model on real ALS data using PRO-ACT dataset. Finally, we compared the progression speed, the estimated reference time, and spatial variability across sex and onset site subgroups in complement to what was done in chapter 2. The Joint cause-specific Spatiotemporal model achieved similar performance to the reference model while capturing an underlying shared latent process, the latent disease age, associated with the impact of the ordering of longitudinal outcomes on the occurrence of the events. This enabled us to overcome the limitation of the reference model that only captures the impact of longitudinal outcomes on events occurrence. The application study gave expected results for the Longitudinal outcomes and showed how to interpret the model.

Contents

4.1 Introduction

Respiratory failure is the leading cause of death in ALS and Non-Invasive Ventilation (NIV) has been proven to be of effective support (Kleopa et al., 1999; Bourke et al., 2006; Hirose et al., 2018). The timing anticipation of its initiation remains a challenge for clinicians due to the heterogeneous clinical manifestation of the disease already described in chapter 2. A better understanding could help improve patient support as its initiation is also symptomatic of an advanced stage of the disease.

Studies report widely different guidelines for the indication of NIV (Georges et al., 2017; Morelot-Panzini, Bruneteau, and Gonzalez-Bermejo, 2019) rather than expected initiation timings. One study (Segura et al., 2023), focusing only on the onset site as a source of heterogeneity, has already demonstrated disparities in NIV initiation. Life support initiations (gastrostomy, tracheostomy, NIV) were also the subject of a work I co-supervised, on the French medico-administrative hospital database, in which both sex and onset site were shown to play a role (Dibling et al., 2024). Nevertheless, longitudinal outcomes of patients were not available for the study. A very recent study analysed the impact of sex and site of onset on NIV initiation (Grassano et al., 2024), but they did not take into account the potential censure of NIV by death, which is known to introduce biases (see section 1.3.1).

As shown in chapter 2 many longitudinal outcomes described the disease progression. Thus, using a joint model can give insight into how their progression is linked to NIV initiation. As this event can be censored by death or tracheostomy, its study, jointly with longitudinal outcomes, requires a joint competing risk model (see section 1.3.3).

To capture individual variability from longitudinal outcomes and model events with competing risks, shared random effects models have been mainly extended using cause-specific (Elashoff, Li, and Li, 2007; Andrinopoulou et al., 2017; Hao et al., 2024) (see section 1.3.3). As described in chapter 3, these models do not realign partial trajectories (Young et al., 2024) and only model the impact of the longitudinal outcome on survival.

To bridge this gap, we extended the Joint Temporal model developed in chapter 3 and studied NIV initiation occurrence jointly with the subscores of ALSFRSr. To do so, we first described the Joint cause-specific Spatiotemporal model: the multivariate Longitudinal Spatiotemporal model associated with a cause-specific Weibull model. We validated it on simulated data and benchmarked it against reference models on PRO-ACT data. Finally, we used it to compare the progression speed, the estimated reference time, and spatial variability across sex and onset site subgroups to complement chapter 2 with death and NIV initiation variability.

4.2 Model specifications

4.2.1 Notations

For the longitudinal process, we used the same notations as the one introduced in section 2.2.1. We considered N patients, indexed by i, followed during n_i visits indexed by j. We modeled K outcomes at each visit so that $y_{i,j,k} = \gamma_{i,k}(t_{i,j,k}) + \epsilon_{i,j,k}$.

For the survival process, we extended the notation defined in section 3.2.1 following the notation of Andrinopoulou et al., 2017. We consider L events associated with one timing t_{e_i} that corresponds to the time of the first event observed, or the censoring time. Then, we associated $B_{e_i} = 0$ if the event is censored and $B_{e_i} = l$ if the event l is observed.

4.2.2 Joint cause-specific Spatiotemporal model

4.2.2.1 Temporal and spatial random effects

Temporal random effects A first individual variability of the patient i is encapsulated in the latent disease age $\psi_i(t) = e^{\xi_i}(t - \tau_i) + t_0$ with e^{ξ_i} is the individual speed factor and τ_i the individual estimated reference time (see section 2.2.3.2). As for the Joint Temporal model, the longitudinal and survival processes $(\gamma_i(t), S_i(t))$ were computed from the latent disease age (see section 3.2.2.1). Thus, to use the model the user must make the hypothesis that the longitudinal and the survival processes are linked by a shared latent disease age, referred to as the shared latent age hypothesis

Spatial random effects A second individual variability, in the disease presentation, is captured as in section 2.2.3.2, by the space shifts $w_i = As_i$ through a dimension reduction with a mixing matrix A and $N_s \leq K-1$ independent sources. As in section 2.2.3.2, to ensure identifiability through Exp-parallelisation, the matrix A is defined as a linear combination of vectors of an orthonormal basis, $(B_k)_{1\leq k\leq K}$, of the hyperplane orthogonal to $Span(v_0)$ (dimension $K \times (K-1)$): each column m of A is thus $A_m = \sum_{o=1}^{K-1} \beta_{o,m} B_o$ with β the matrix of coefficient (dimension $(K-1) \times N_s$) so that $A = (B\beta)^T$. These sources are also used to link the survival and the longitudinal process with the creation, for each event l, of a survival shift $u_{i,l} = \sum_{m=1}^{N_s} \zeta_{l,m} s_{i,m}$ with ζ a matrix of hazard ratio coefficients. Note that, to describe individual variability on one longitudinal outcome k or event l, space shift $w_{i,k}$, and survival shits $u_{i,l}$ are usually easier to interpret compared to sources s_i, as they encapsulate the total effect of the spatial variability.

4.2.2.2 Longitudinal submodel

Longitudinal process modelling consists of calculating logistic trajectories $\gamma_{i,k}(t)$ for each outcome k, from the latent disease age $\psi_i(t)$, taking into account space shifts $w_{i,k}$ (see section 2.2.3.3):

$$\gamma_{i,k}(t) = \left(1 + g_k \times \exp(-\tfrac{(1+g_k)^2}{g_k}(v_{0,k}(\psi_i(t) - t_0) + w_{i,k})\right)^{-1} \tag{4.1}$$

where t_0 is the population estimated reference time, $v_{0,k}$ is the speed of the logistic curve at t_0 and $\frac{1}{1+g_k}$ is the value of the logistic curve at t_0.

4.2.2.3 Survival submodel

We used a cause-specific structure (Prentice and Gloeckler, 1978; Cheng, Fine, and Wei, 1998) to handle competing risks. Doing so, for each event l and patient i, we define a hazard $h_{i,l}(t)$, and an associated survival function $S_{i,l}(t)$. The Weibull distribution is used to model

the event occurrence from the latent disease age with an additional Cox-proportional hazard impact of the sources on the hazard using the survival shifts $u_i = \zeta s_i$:

$$\begin{aligned} h_{i,l}(t) = & \quad h_{0,i,l}(t) \exp(u_{i,l}) \\ = & \quad \frac{\rho_l e^{\xi_i}}{\nu_l} \left(\frac{e^{\xi_i}(t-\tau_i)}{\nu_l} \right)^{\rho_l - 1} \exp(u_{i,l}) \end{aligned}$$

where ν_l is the scale and ρ_l the shape of the Weibull distribution of the event l. From there, we compute the survival of event l:

$$\begin{aligned} S_{i,l}(t) = & \quad \exp\left(- \int_0^t h_{i,l}(x) dx \right) \\ = & \quad \exp\left(- \left(\frac{e^{\xi_i}(t-\tau_i)}{\nu_l} \right)^{\rho_l} \exp(u_{i,l}) \right) \end{aligned}$$

And the Cumulative Incidence Function (CIF) of event l:

$$\begin{aligned} CIF_{i,l}(t) = & \quad \int_0^t h_{i,l}(x) \prod_q^L S_{i,q}(x) dx \\ = & \quad \int_0^t \frac{\rho_l e^{\xi_i}}{\nu_l} \left(\frac{e^{\xi_i}(x-\tau_i)}{\nu_l} \right)^{\rho_l - 1} \exp(u_{i,l}) \prod_q^L \exp\left(- \left(\frac{e^{\xi_i}(x-\tau_i)}{\nu_q} \right)^{\rho_q} \exp(u_{i,q}) \right) dx \end{aligned}$$

We can see here that the survival shits $u_{i,l}$ have a proportional effect on the hazard and thus have an interpretation close to the classic Hazard Ratio. Nevertheless, as they are not exactly the same, for clarity reasons they will be referred to as the Proportional effect of survival shits on the Hazard (PH).

4.2.2.4 Joint cause-specific Spatiotemporal model

The Joint cause-specific Spatiotemporal model can thus be summarised for a patient i, an outcome k, and an event l by:

$$\begin{cases} \psi_i(t) = e^{\xi_i}(t - \tau_i) + t_0 \\ w_i = A s_i \\ u_i = \zeta s_i \\ \gamma_{i,k}(t) = \left(1 + g_k \times \exp(-v_{0,k} \frac{(g_k+1)^2}{g_k} e^{\xi_i}(t - \tau_i) + w_{i,k}) \right)^{-1} \\ S_{i,l}(t) = \exp\left(- \left(\frac{e^{\xi_i}(t-\tau_i)}{\nu_l} \right)^{\rho_l} \exp(u_{i,l}) \right) \end{cases} \tag{4.2}$$

4.2.3 Likelihood of the Joint cause-specific Spatiotemporal model

4.2.3.1 Parameters

We put ourselves in the same context as in section 2.2.4.1. The parameters of the model can be summarised as follows with the patients indexed by i and outcomes by k, the events by l, the sources by m, and the dimensions of the hyperplane orthogonal to $Span(v_0)$ by o:

- Latent parameters (z):

– Latent fixed effects (z_{fe}): fixed effects sampled

$$\tilde{g}_k = \log(g_k) \sim \mathcal{N}\left(\overline{\tilde{g}_k}, \sigma^2_{\tilde{g}}\right) \qquad \tilde{v}_{0,k} = \log(v_{0,k}) \sim \mathcal{N}\left(\overline{\tilde{v}_{0,k}}, \sigma^2_{\tilde{v}_0}\right)$$
$$\tilde{\nu}_l = -\log(\nu_l) \sim \mathcal{N}\left(\overline{\tilde{\nu}_l}, \sigma^2_{\tilde{\nu}}\right) \qquad \tilde{\rho}_l = \log(\rho_l) \sim \mathcal{N}\left(\overline{\tilde{\rho}_l}, \sigma^2_{\tilde{\rho}}\right)$$
$$\zeta_{l,m} \sim \mathcal{N}\left(\overline{\zeta}_{l,m}, \sigma^2_{\zeta}\right) \qquad \beta_{o,m} \sim \mathcal{N}\left(\overline{\beta}_{o,m}, \sigma^2_{\beta}\right)$$

– Latent random effects (z_{re}): random effects sampled

$$\xi_i \sim \mathcal{N}\left(\overline{\xi}, \sigma^2_{\xi}\right) \qquad \tau_i \sim \mathcal{N}\left(\overline{\tau}, \sigma^2_{\tau}\right) \qquad s_{i,m} \sim \mathcal{N}\left(\overline{s}, \sigma_s\right)$$

- Model parameters (θ): fixed effects estimated from log-likelihood maximisation $\theta = \{\overline{\tilde{g}_k}, \overline{\tilde{v}_{0,k}}, \overline{\tilde{\nu}_l}, \overline{\tilde{\rho}_l}, \overline{\beta_{o,m}}, \overline{\zeta_{l,m}}, \sigma, \sigma_\xi, t_0, \sigma_\tau\}$
- Hyperparameters (Π): set by the user $\Pi = \{\sigma_{\tilde{g}}, \sigma_{\tilde{v}_0}, \sigma_{\tilde{\nu}}, \sigma_{\tilde{\rho}}, \sigma_\beta, \sigma_\zeta, \sigma_s\}$

To ensure identifiability, we set $\overline{\xi} = 0$, $\sigma_s = 1$, $\overline{s} = 0$ and $t_0 = \overline{\tau}$.

4.2.3.2 Log-likelihood structure

The likelihood estimated by the model is the following:

$$p(y, T_e, B_e \mid \theta, \Pi) = \int_z p(y, T_e, B_e, z \mid \theta, \Pi) dz$$

As in chapter 3.2.3.2, the $p(y, T_e, B_e, z \mid \theta, \Pi)$ could be divided into four different terms: the longitudinal data attachment, the survival data attachment, two terms for the prior attachment of latent parameters (fixed and random). We end up with the following expression :

$$\begin{aligned}\log p((y, t_e, B_e), z \mid \theta, \Pi) = \quad & \log p(y \mid z, \theta, \Pi) \\ + \quad & \log p(t_e, B_e \mid z, \theta, \Pi) \\ + \quad & \log p(z_{re} \mid z_{fe}, \theta, \Pi) \\ + \quad & \log p(z_{fe} \mid \theta, \Pi)\end{aligned}$$

The different parts of the log-likelihood are described below associated with their different assumptions, with the priors attachment to latent fixed effect, $p(z_{fe} \mid \theta, \Pi)$, separated for longitudinal and survival effects. Note that the total formula of the full log-likelihood is available in appendix A.1.

Longitudinal data attachment The longitudinal data attachment, $p(y \mid z, \theta, \Pi)$, is the same, with the same assumptions, as the one from the Longitudinal Spatiotemporal model (see section 2.2.4.2).

Survival data attachment To model the survival process, we assumed that all patients were independent and that the modelling of the survival process depended on whether the event was observed or not. Note that the following equation could be interpreted as follows:

the patient must have survived till the time of observation (or censure) and then has an instantaneous risk for the observed events (Mozumder, Rutherford, and Lambert, 2018):

$$\begin{aligned}
\log p(t_e, B_e \mid z, \theta, \Pi) = & \sum_i \log p(t_{e_i}, B_{e_i} \mid z, \theta, \Pi) \\
= & \sum_{i,l} \mathbb{1}_{B_{e_i}=l} \times \log\left(h_{i,l}(t_{e_i})\right) + \sum_{i,l} \log\left(S_{i,l}(t_{e_i})\right) \\
= & \sum_{i,l} \mathbb{1}_{B_{e_i}=l} \times \log\left(\frac{\rho_l e^{\xi_i}}{\nu_l}\left(\frac{e^{\xi_i}(t_{e_i}-\tau_i)}{\nu_l}\right)^{\rho_l - 1} \exp\left(u_{i,l}\right)\right) \\
- & \sum_{i,l} \left(\frac{e^{\xi_i}(t_{e_i}-\tau_i)}{\nu_l}\right)^{\rho_l} \exp\left(u_{i,l}\right)
\end{aligned}$$

Note that the likelihood for a simple event could be extracted from the above formula by putting $L = 1$. As in section 3.2.3.2, if $\psi_i(t) < t_0$ $\log\left(h_i(t_{e_i})\right) = -\infty$. To prevent estimation issues, we initialised the algorithm at a possible point getting inspiration from barrier methods (Nesterov, 2018).

Latent random effects priors attachment The latent random effects priors attachment, $p(z_{re} \mid z_{fe}, \theta, \Pi)$, is the same, with the same assumptions, as the one from the Longitudinal Spatiotemporal model (see section 2.2.4.2).

Latent fixed effects priors longitudinal attachment The latent fixed effects priors longitudinal attachment $p(\tilde{g}, \tilde{v}_0, \beta \mid \theta, \Pi)$ is the same, with the same assumptions, as the fixed effects priors attachment, $p(z_{fe} \mid \theta, \Pi)$, from the Longitudinal Spatiotemporal model (see section 2.2.4.2).

Latent fixed effects prior event attachment Each latent fixed effect is independently sampled from a posterior distribution. The regularization term associated is then:

$$\begin{aligned}
\log p(\tilde{\nu}, \tilde{\rho}, \zeta \mid \theta, \Pi) = & \sum_l \log p(\tilde{\nu}_l \mid \theta, \Pi) + \log p(\tilde{\rho}_l \mid \theta, \Pi) \\
+ & \sum_{l,m} \log p(\zeta_{l,m} \mid \theta, \Pi) \\
= - & \sum_l \log\left(\sigma_{\tilde{\nu}_l}\sqrt{2\pi}\right) - \frac{1}{2\sigma^2_{\tilde{\nu}_l}}\left(\tilde{\nu}_l - \overline{\tilde{\nu}_l}\right)^2 \\
- & \sum_l \log\left(\sigma_{\tilde{\rho}_l}\sqrt{2\pi}\right) - \frac{1}{2\sigma^2_{\tilde{\rho}_l}}\left(\tilde{\rho}_l - \overline{\tilde{\rho}_l}\right)^2 \\
- & \sum_{l,m} \log\left(\sigma_\zeta\sqrt{2\pi}\right) - \frac{1}{2\sigma^2_\zeta}\left(\zeta_{l,m} - \overline{\zeta}_{l,m}\right)^2
\end{aligned}$$

4.2.4 Estimation of the Joint cause-specific Spatiotemporal model

The estimation method can be separated into two steps, as in section 2.2.5. The calibration is the same as the one used in section 2.2.5. The adapted sufficient statistics and maximisation update rules, necessary for the computation, are given in the appendix A.2 and A.3.

For the personalisation, the only difference with section 2.2.5 is, that the CIF is needed to compute the predictions. The CIF is corrected using the survival probability at the last visit used for personnalisation as in Andrinopoulou et al., 2017.

An implementation of the Joint cause-specific Spatiotemporal model is available in the open-source package Leaspy (v2): `https://gitlab.com/icm-institute/aramislab/leaspy`.

4.2.5 Reference multivariate and cause-specific models

Several reference models were used in this chapter. First, we used one-process-only models. For the survival model, we used a cause-specific Weibull AFT model to describe the survival process, using the R flexsurv package (Jackson, 2016). This model will be referred to as the cause-specific AFT model. For the longitudinal model, we used the Longitudinal Spatiotemporal model described in section 3.2.2.2 using Leaspy package. This model will be referred to as the Longitudinal model.

Then, we used a joint model with shared random effects for competing risks. We used a logistic longitudinal process, with a cause-specific competing risk model, estimated using the JMbayes2 package (Rizopoulos, 2016). This model will be referred to as the cause-specific JMbayes2 model. All the model equations are summarised in Table 4.1.

TABLE 4.1: Specification of reference models for multivariate longitudinal outcomes and competing risks

Legend: Longitudinal: Longitudinal Spatiotemporal model, Cause-specific AFT: Cause-specific Weibull Accelerated Failure Time model, Joint Spatiotemporal: the Joint cause-specific Spatiotemporal model, Cause-specific-JMbayes2: joint model with shared random effects and cause-specific survival model. spl(t): spline function. L events indexed by l and K longitudinal outcomes indexed by k. For space reasons, the CIF is not integrated into the table, but all the model followed a cause-specific structure described in Zhang, Zhang, and Scheike, 2008 and in section 4.2.2.3

Model	Inputs	Effects		Random effects structure	Link functions	
		Fixed	Random	$\psi_i(t)$	Longitudinal $\gamma_{i,k}(t)$	$S_{i,l}(t)$
Longitudinal	t	g, v_0, t_0	ξ_i, τ_i, w_i	$e^{\xi_i}(t-\tau_i)+t_0$	$\left(1+g_k \times \exp(-v_{0,k}\frac{(g_k+1)^2}{g_k}(\psi_i(t)-t_0)+w_{i,k})\right)^{-1}$	-
Cause-specific AFT	t	ν_0, ρ_0	-	-	-	$exp\left(-\left(\frac{t}{\nu_l}\right)^{\rho_l}\right)$
Joint Spatiotemporal	t	g, v_0, t_0, ρ, ν	ξ_i, τ_i, w_i, u_i	$e^{\xi_i}(t-\tau_i)+t_0$	$\left(1+g_k \times \exp(-v_{0,k}\frac{(g_k+1)^2}{g_k}(\psi_i(t)-t_0)+w_{i,k})\right)^{-1}$	$\exp\left(-\left(\frac{\psi_i(t)}{\nu_l}\right)^{\rho_l}\exp(u_{i,l})\right)$
Cause-specific-JMBayes2	t	$\beta_0, \beta_1, g_k, \alpha, spl(t)$	$b_{i,0}, b_{i,1}$	$(\beta_0+b_{i,0})+(\beta_1+b_{i,1})t$	$(1+g_k \times \exp(\psi_i(t)))^{-1}$	$\exp\left(-\int_0^t \exp(spl_l(u)+\sum_k \alpha_k \gamma_{i,k}(u))du\right)$

4.3 Material

4.3.1 PRO-ACT data

4.3.1.1 Population

We extracted patients from the PRO-ACT database with age at first symptoms, sex, and onset site. To limit left-censored VNI initiation, we selected patients with a Mitos score equal to 0 (Fang et al., 2017). We used this population to extract real-like values for the simulation study and perform the application study. For the benchmark study, we sub-selected this population by keeping only patients with at least three visits to be able to evaluate longitudinal predictions: on the test set, we personalised the models on two visits and made predictions on the remaining.

4.3.1.2 Outcome

Longitudinal outcomes For the benchmark and the application, we used three subscores of ALSFRSr: bulbar scale, fine motor scale, and gross motor scale (Rooney et al., 2017). Indeed, we did not want to use respiratory longitudinal outcomes that risk to capture all the correlations with events.

To test the impact of the number of sources (at a maximum of K-1, see section 2.2.3.2), we wanted to simulate four longitudinal outcomes. We added the ALSFRSr total, which does not make much sense from a clinical point of view, but the idea was only to have credible parameters for the simulation of four outcomes. We normalized the outcomes between 0 (the healthiest value) and +1 (the maximum pathological change). All the scores were normalized using their theoretical maximum and minimum values (Rooney et al., 2017). We reindexed the visits by the time from symptom onset.

Survival outcomes We extracted the death and NIV initiation age from the PRO-ACT database as described in section 5.3.3. As NIV initiation was interval censored, we used the mean of the interval as an approximation, even though this might introduce some biases (Leffondré et al., 2013). Death and tracheostomy were also extracted and associated as in the majority of ALS studies. Note that for simplicity, we will talk about death to encapsulate both, in the following sections. As visits, events were reindexed by the time from symptom onset.

4.3.2 Simulated data

4.3.2.1 Data simulation method

Data were simulated under our Joint cause-specific Spatiotemporal model structure. Parameter values for each scenario are summarised in Table 4.2. We simulated data thanks to the following procedure:

1. We simulated random effects using $\xi_i \sim \mathcal{N}\left(0, \sigma_\xi^2\right)$, $\tau_i \sim \mathcal{N}\left(t_0, \sigma_\tau^2\right)$, and N_s sources $s_{i,m} \sim \mathcal{N}\left(\overline{s}, \sigma_s\right)$.

2. We modelled age at first visit (baseline) $t_{b,i}$ as $t_{b,i} = \tau_i + \delta_{b_i}$ with $\delta_{b_i} \sim \mathcal{N}\left(\overline{\delta_b}, \sigma^2_{\delta_b}\right)$.

3. We set a time of follow-up per patient T_{f_i}, with $T_{f_i} \sim \mathcal{N}\left(\overline{T_f}, \sigma^2_{T_f}\right)$ and a time between two visits $\delta_{v_{i,j}} = t_{i,j-1} - t_{i,j}$, with $\delta_{v_i} \sim \mathcal{N}\left(\overline{\delta_v}, \sigma^2_{\delta_v}\right)$ to simulate n_i visits until $t_{i,n_i} \leq t_{i,0} + T_{f_i} < t_{i,n_{i+1}}$.

4. We set the value of the K longitudinal outcomes at each visit using $y_{i,j,k} = \gamma_{i,k}(t_{i,j,k}) + \epsilon_{i,j,k}$ with $\epsilon_{i,j,k} \sim \mathcal{N}\left(0, \sigma^2_k\right)$ with Leaspy software.

5. For each patient, we simulated the L event $T_{e_{i,l}}$ using $T_{e_{i,l}} \sim e^{-\xi_i}\mathcal{W}\left(\nu_l, \rho_l\right) + \tau_i$.

6. We kept the first event that occurred as observed and censored the others,

7. We considered that the first event stopped the follow-up and that the follow-up censored the first event. Thus all the visits of each outcome k after the event were censored: $t_{i,j,k} > T_{e_i}$, and events after the last visit were censored: $max(t_{i,j,k}) < T_{e_i}$.

As we studied sub-cores of one score, we considered that all measures were available at a given time.

4.3.2.2 Scenarios

As in section 3.3.1.2, we used the PRO-ACT dataset, described in section 2.3.1, to get estimated real-like values for parameters. Parameters directly associated with the disease were extracted from data analysis, using the Longitudinal and cause-specific Weibull AFT models.

To validate the model and give future users insight on how to select the right number of sources, we simulated four outcomes with two sources, to be able to evaluate the model with 1, 2, and 3 sources. We simulated a dataset close to real data with a valid shared latent age hypothesis, following the procedure of section 4.3.2.1. This dataset will be referred to as the Real-like dataset.

We also simulated longitudinal and survival data with no link, to give insight on how to test if the shared latent age hypothesis is valid. To do so, we simulated two sets of random effects, one for the longitudinal and one for the survival process. Apart from that, we kept the same values and process for simulation. This dataset will be referred to as the no-link dataset.

TABLE 4.2: Data simulation parameters by scenario

Legend: Real-like: Real-like simulated dataset (valid shared latent age hypothesis), No link: No-link dataset (invalid shared latent age hypothesis), (r) indicate when ALS real-like parameters are used. More information on simulation is available in section 4.3.2. $\overline{\tau} = t_0, \overline{\xi} = 0$, $\sigma_s = 1$ *and* $\overline{s} = 0$

Type	Parameters Name		ALS Symbol	Real-like parameters (r)	Real-like simulation	No-link simulation
Patients	Patient number		N	1,000	r	r
Random Effect	Population estimated reference time (year)	(mean)	$\overline{\tau}$	1.3	r	r
		(std)	σ_τ	1.1	r	r
	Individual log-speed factor	(mean)	$\overline{\xi}$	0	r	r
		(std)	σ_ξ	0.79	r	r
	Number of sources		N_s	2	r	r
Longitudinal Fixed Effects	Speed of the logistic curve		v_0	[0.07 , 0.19, 0.20, 0.11]	r	r
	Curve value at t_0: $\frac{1}{1+g}$		g	[13.96, 5.32, 3.99, 5.70]	r	r
	Estimated noise		σ	[0.07, 0.08, 0.07, 0.04]	r	r
	Longitudinal mixing matrix		A	[[0.059, -0.103, 0.001, 0.004], [0.059, 0.006, -0.141, -0.004]]	[[0.06, -0.10, 0.00, 0.01], [0.06, 0.006, -0.14, -0.00]]	[[0.06, -0.10, 0.00, 0.01], [0.06, 0.01, -0.14, -0.00]]
Survival Fixed Effects	Scale of the Weibull distribution		ν	[3.4, 3.9]	[2.8, 3.6]	[2.8, 3.6]
	Shape of the Weibull distribution		ρ	[1.7, 2.8]	r	r
	Hazard ratio coefficients		ζ	[[-0.09, 0.09] [-0.1, 0.05]]	[[-0.09, 0.09] [-0.1, 0.0]]	[[-0.09, 0.09] [-0.1, 0.0]]
Visits	Time between τ and baseline (years)	(mean)	$\overline{\delta_b}$	0.2	0.3	0.3
		(std)	σ_{δ_b}	0.8	r	r
	Time of follow up (year)	(mean)	$\overline{T_f}$	0.8	1.1	1.1
		(std)	σ_{T_f}	0.5	r	r
	Time between visits (months)	(mean)	$\overline{\delta_v}$	1.4	2.0	2.0
		(std)	σ_{δ_v}	0.75	r	r

4.3.3 Characteristics of the datasets

Out of the 8,571 patients from the PRO-ACT database, we subselected 6,034 patients with sex and first symptoms (spinal or bulbar onset) provided. Out of them, 2,219 had their first visit with a Mitos score equal to 0. Then 42 patients were dropped for the Analysis dataset due to left censored VNI. For the Benchmark dataset, we also dropped patients with less than 3 visits and ended up with 1,919 patients. Characteristics of the Analysis and the Benchmark dataset were close despite the subselection of patients (Table 4.3). NIV initiation was interval censored between two visits, with a mean interval of 1.9 (1.4) months.

Simulated scenarios had fewer patients and visits than the PRO-ACT datasets, even though the time of follow-up was a little bit longer (Table 4.3). We also added slightly more censoring of events which also increased the modeling difficulty. Due to the censure of the fastest patients in the simulation process (some events happened before their first visit), the final simulated value of the Weibull scale, ν, and the population speed, v_0, were respectively higher and smaller compared to the target simulation.

TABLE 4.3: Characteristics of the PRO-ACT and simulated datasets

Legend: Results are presented with mean (SD) [class%]. There were no missing values in the dataset due to patient selection.
Analysis: extraction from the PRO-ACT database used for the application and the estimation of parameters used for simulation, Benchmark: extraction from PRO-ACT database used to benchmark the models, patients have at least 3 visits, Real-like: Real-like simulated dataset (valid shared latent age hypothesis), No link: No-link dataset (invalid shared latent age hypothesis).

		Real PRO-ACT data		Simulated data	
Type	Characteristics	Analysis	Benchmark	Real-like	No link
Number	patients	2,177	1,919	1,000	1,000
	visits	16,400	16,036	6,401	6,879
	patient-years	1,661	1,650	901	973
	visits per patients	7.5 (4.5)	8.4 (4.1)	6.4 (3.2)	6.9 (3.4)
Time	follow-up (years)	0.8 (0.5)	0.9 (0.5)	0.9 (0.5)	1.0 (0.6)
	between visits (months)	1.4 (0.7)	1.4 (0.7)	2.0 (0.7)	2.0 (0.7)
Gender	(Male)	1,364 (62.7 %)	1197 (62.4 %)	-	-
Symptom onset	(Spinal)	1,666 (76.5 %)	1465 (76.3 %)	-	-
Age at first symptoms		54.1 (11.3)	54.0 ± 11.4	-	-
Observed events (%)	VNI	570 (26.2%)	477 (24.9%)	204 (20.4 %)	171 (17.1%)
	Death	245 (11.3%)	216 (11.3%)	91 (9.1%)	56 (5.6%)
ALSFRSr (baseline)	total	39.4 (4.1)	39.6 (4.1)	38.8 (5.3)	36.7 (8.6)
	bulbar	10.3 (2.0)	10.3 (2.0)	10.2 (2.0)	9.7 (2.7)
	fine motor	9.1 (2.0)	9.1 (2.0)	8.9 (2.6)	8.3 (3.2)
	gross motor	8.5 (2.4)	8.6 (2.5)	8.5 (2.7)	7.8 (3.2)

4.4 Simulation study

The objective of this section was to validate the Joint cause-specific Spatiotemporal model by assessing fixed and random effects estimation on simulated data and give insight into how to test the shared latent age hypothesis.

4.4.1 Method

For each experiment, we first trained the Longitudinal model during 2,000 iterations. Then, we ran the Joint cause-specific Spatiotemporal model for 70,000 iterations (on average 7 hours) using the values of the Longitudinal model as initialisation and with a Robbins-Monroe convergence phase on the 10,000 last iterations (Robbins and Monro, 1951). The different parameters estimated by the Joint cause-specific Spatiotemporal model for the simulation study are summarised in Table 4.4 jointly with their interpretation.

TABLE 4.4: Description of the parameters estimated by the Joint cause-specific Spatiotemporal model for the simulation study

Legend: m indexed the dimension of the ordering of longitudinal outcomes (spatial variability), here with dimension three (three sources). Mixing matrix and HR coefficients must be adaptated when using one (first line only) or two sources (two first lines). Individual log-speed factor mean $\overline{\xi}$, individual spatial variability mean and standard deviation $\overline{s}, \sigma_s$ parameters are not present as they are fixed by the model ($\overline{\xi} = 0, \overline{s} = 0, \sigma_s = 1$) and $t_0 = \overline{\tau}$

Distribution of random effects				
Estimated reference time (mean)		t_0		
Estimated reference time (std)		σ_τ		
Individual log-speed factor (std)		σ_ξ		
Longitudinal fixed effects				
	Bulbar	Fine motor	Gros motor	Total
Curve values at t_0: $\frac{1}{1+g_k}$ (g_k)	g_0	g_1	g_2	g_3
Speed of the logistic curves $(v_{0,k})$	$v_{0,0}$	$v_{0,1}$	$v_{0,2}$	$v_{0,1}$
Estimated noises (σ_k)	σ_0	σ_1	σ_2	σ_3
Mixing matrix $(A_{k,m})$	$A_{0,0}$	$A_{1,0}$	$A_{2,0}$	$A_{3,0}$
	$A_{0,1}$	$A_{1,1}$	$A_{2,1}$	$A_{3,1}$
	$A_{0,2}$	$A_{1,2}$	$A_{2,2}$	$A_{3,2}$
Survival fixed effects				
		NIV initiation	Death	
Weibull scale (ν_l)		ν_0	ν_1	
Weibull shape (ρ_l)		ρ_0	ρ_1	
HR Coefficients $(\zeta_{l,m})$		$\zeta_{0,0}$	$\zeta_{1,0}$	
		$\zeta_{0,1}$	$\zeta_{1,1}$	
		$\zeta_{0,2}$	$\zeta_{1,2}$	

4.4.1.1 Determining the number of sources

The number of sources is a hyperparameter that we can adjust. It corresponds to the number of dimensions allowed for the ordering of the longitudinal outcomes (spatial aspect). To evaluate the optimal number of sources needed to model the data, as we got four longitudinal outcomes to model, we ran three models with 1, 2, and 3 sources on the datasets. Indeed the number of sources must be below or equal to the number of longitudinal outcomes

minus one (see section 4.2.2.1). To compare the models, we used the conditional version of the Watanabe–Akaike Information Criterion (WAIC) (Watanabe, 2010) (see appendix C.1), a generalised version of the AIC, which gives a measure of the trade-off between the goodness of fit and the simplicity of the model. We then kept the number of sources that enabled to minimise the conditional WAIC.

4.4.1.2 Precision of the parameters estimated

To validate the model, we assessed how well the model could estimate fixed and random effects when the shared latent age hypothesis (the longitudinal and the survival processes are linked by a shared latent disease age (see section 4.2.2.1)) is valid (Real-like simulated dataset) and not valid (No-link simulated dataset). For fixed effects, as done by Pan and Hout, 2023, we extracted data from the posterior distribution (between the 50.000 and 60.000 iterations). We obtained the 95% credibility interval of the estimated fixed effects and checked if the simulated value was included in the credibility interval. For random effects, we extracted the mean of the random effects posteriors for each individual (between 50,000 and 60,000 iterations) to reduce the computation complexity. As a metric, we used the intraclass correlation between the value obtained with the simulated dataset for each individual and the true value that enabled the simulation.

4.4.1.3 Validity of the shared latent age hypothesis

Finally, we wanted to give future users tools to evaluate if the shared latent age hypothesis was realistic on their dataset.

To do so, we compared the estimated parameters from the Joint cause-specific Spatiotemporal model with the one of the cause-specific AFT Weibull model and the Longitudinal Spatiotemporal model, both on the Real-like and No-link simulated dataset. For the cause-specific AFT Weibull model we matched ν with $t_0 + \nu$ in the Joint cause-specific Spatiotemporal model, as the survival submodel starts from t_0 (see section 4.2).

4.4.2 Results

4.4.2.1 Determining the number of sources

The objective was to evaluate whether the conditional WAIC could be used to select the correct value for the number of sources of the model, the only hyperparameter that needs to be set by the user.

On the Real-like dataset, the conditional WAIC was lower for the model with two sources (-73,628) compared to the one with one (-59,969) and three (-71,211) sources. These results confirmed what was expected as we simulated data with two sources.

On the No-link dataset, the conditional WAIC was lower for the model with three sources (-79,631) compared to the one with one (-65,179) and two (-79,144) sources. The data were simulated using two sources, so the test failed to detect it. Nevertheless, this might be linked to the fact that the shared latent age hypothesis was not valid for the No-link dataset.

We validated the use of the conditional WAIC to select the number of sources when the shared latent age hypothesis was valid. In the following parts, a Joint cause-specific model with two sources on the Real-like dataset was used, and with three sources on the No-link dataset.

4.4.2.2 Precision of the parameters estimated

We wanted to evaluate the performance of the model on simulated data with valid (Real-like dataset) and invalid (No-link dataset) shared latent age hypotheses.

Real-like dataset On the Real-like dataset, out of 32, 18 of the 95% credibility interval of the distribution contained the simulated parameters (Table 4.5). For the simulated parameters that did not fall in the 95% credibility interval of the distribution, the mean absolute bias was below 28% of the simulated value. We can note, that as in chapter 3, population speed, v_0, tended to be underestimated and Weibull scale, ν, overestimated, but biases were relatively small.

The hazard ratio coefficients (ζ) estimate the association between the ordering of the longitudinal outcomes and the events occurring. The value of $\zeta_{1,1}$ was found not significantly different from zero which corresponded to the simulated value. Nevertheless, $\zeta_{0,0}$ was also found not significantly different from zero even though it was simulated with -0.09. This might be linked to the small size of the value.

Both temporal and spatial random effects were well reconstructed, with intra-class correlations higher than 0.83 (Table C.1 in appendix).

No-link dataset Here the idea is mainly to test what happens when this hypothesis is invalid. On the No-link dataset, out of 37, 11 of the 95% credibility interval of the distribution contained the simulated parameters (Table C.2 in appendix). The parameters linked to the distribution of the random effects and the longitudinal fixed effects, except the values of the mixing matrix, were close to the values of the parameters used for simulation, with less than 15% of bias (Table C.2 in appendix). The mixing matrix and the survival fixed effects were far from what was simulated going for the most extreme values to 1,000% of bias (Table C.2 in appendix).

The hazard ratio coefficients (ζ) failed to capture the absence of association between the ordering of the longitudinal outcomes and the events occurring: 3 out of 6 coefficients were found significantly different from zero. This might be explained by the fact that the shared latent age hypothesis was not valid on the No-link dataset.

Temporal random effects and space shifts were well reconstructed for the longitudinal process (intra-class correlations higher than 0.9) but not for the survival shifts ($u_{i,1}$: -0.91 [-0.92 -0.9] and $u_{i,2}$: 0.01 [-0.05 0.07]) (Table C.3 in appendix).

When the shared latent age hypothesis was valid, the model well reconstructs the data with 18, out of 32, of the 95% credibility interval of the distribution that contained the simulated parameters and mean absolute biases lower than 30% otherwise. Nevertheless, when the shared latent age hypothesis was invalid, the model only fits the longitudinal

process, resulting in wrong survival model fixed effects. Note that this behavior might be due to the quantity of information available in the data. Such a situation must be diagnosed by a user and the following section proposes a methodology to do so.

TABLE 4.5: Biases of the estimated parameters of the Joint cause-specific Spatiotemporal model on the Real-like simulated dataset

Legend: For each parameter of the model, the value used for simulation is reported with the value of the mean of the stabilized posterior and the associated Credibility Interval at 95% computed on the last 10,000 iterations of the stabilized MCMC-SAEM: (*) the simulated value is in the 95% credibility interval, (<) the simulated value is underestimated, (>) the simulated value is overestimated. Bias: difference to the true parameters divided by the true parameters,-: put when the bias could not be computed as the true parameter equal to 0, $\overline{\xi}, \overline{s}, \sigma_s$ *parameters are not present as they are fixed by the model (*$\overline{\xi} = 0, \overline{s} = 0, \sigma_s = 1$*) and* $t_0 = \overline{\tau}$

	Parameters name		Simulated		Estimated	Bias (%)
Distribution of random effects	Estimated reference time (mean)	t_0	1.36	*	1.381 [1.336, 1.421]	1.5 [-1.8, 4.4]
	Estimated reference time (std)	σ_τ	1.062	*	1.058 [1.038, 1.080]	-0.4 [-2.3, 1.7]
	Individual log-speed factor (std)	σ_ξ	0.725	*	0.749 [0.722, 0.780]	3.4 [-0.3, 7.6]
Longitudinal fixed effects	Curve values at t_0: $\frac{1}{1+g_k}$ (g_k)	g_0	13.958	<	13.002 [12.344, 13.712]	-6.9 [-11.6, -1.8]
		g_1	5.316	<	4.549 [4.229, 4.850]	-14.4 [-20.4, -8.8]
		g_2	3.993	<	3.639 [3.485, 3.828]	-8.9 [-12.7, -4.1]
		g_3	5.704	<	5.428 [5.308, 5.534]	-4.8 [-6.9, -3.0]
	Speed of the logistic curves ($v_{0,k}$)	$v_{0,0}$	0.062	<	0.059 [0.056, 0.061]	-4.7 [-9.2, -0.3]
		$v_{0,1}$	0.167	*	0.165 [0.157, 0.173]	-1.3 [-6.0, 3.4]
		$v_{0,2}$	0.176	<	0.167 [0.161, 0.173]	-5.0 [-8.6, -1.6]
		$v_{0,3}$	0.1	<	0.094 [0.091, 0.096]	-6.3 [-8.9, -3.8]
	Mixing matrix ($A_{k,m}$)	$A_{0,0}$	0.06	*	0.059 [0.053, 0.067]	-1.5 [-10.8, 12.1]
		$A_{0,1}$	0.059	*	0.059 [0.051, 0.066]	0.8 [-14.0, 12.6]
		$A_{1,0}$	-0.1	>	-0.107 [-0.115, -0.101]	-7.4 [-15.2, -1.1]
		$A_{1,1}$	0.006	*	0.009 [-0.003, 0.023]	50.7 [-146.1, 286.7]
		$A_{2,0}$	0.0	*	-0.005 [-0.024, 0.007]	-
		$A_{2,1}$	-0.14	*	-0.147 [-0.154, -0.138]	-4.9 [-10.1, 1.5]
		$A_{3,0}$	0.01	*	0.01 [0.009, 0.011]	1.0 [-12.0, 14.5]
		$A_{3,1}$	-0.0	*	-0.0 [-0.002, 0.001]	-
	Estimated noises (σ_k)	σ_0	0.065	<	0.058 [0.057, 0.058]	-11.0 [-11.7, -10.3]
		σ_1	0.076	<	0.073 [0.072, 0.074]	-4.0 [-4.8, -3.2]
		σ_2	0.074	<	0.072 [0.071, 0.073]	-3.0 [-3.8, -2.2]
		σ_3	0.036	*	0.036 [0.036, 0.037]	0.1 [-0.7, 1.0]
Survival fixed effects	Weibull scale (ν_l)	ν_0	3.143	>	4.012 [3.706, 4.303]	27.7 [17.9, 36.9]
		ν_1	4.041	>	4.499 [4.226, 4.781]	11.4 [4.6, 18.3]
	Weibull shape (ρ_l)	ρ_0	1.7	*	1.739 [1.618, 1.887]	2.3 [-4.8, 11.0]
		ρ_1	2.8	*	2.908 [2.678, 3.145]	3.9 [-4.3, 12.3]
	HR Coefficients ($\zeta_{l,m}$)	$\zeta_{0,0}$	-0.09	*	-0.071 [-0.161, 0.028]	21.7 [-78.4, 131.6]
		$\zeta_{0,1}$	-0.1	*	-0.076 [-0.212, 0.046]	24.1 [-112.4, 146.3]
		$\zeta_{1,0}$	0.09	*	0.153 [0.006, 0.285]	70.3 [-92.8, 216.4]
		$\zeta_{1,1}$	0.0	*	-0.01 [-0.210, 0.151]	-

4.4.2.3 Validity of the Shared latent age hypothesis

As seen above, it is important to diagnose cases where the shared latent age hypothesis might be invalid. We proposed to compare the results of the Joint Spatiotemporal model

with one process model only, to test such a hypothesis.

Longitudinal submodel On the Real-like datasets, parameters of the Longitudinal Spatiotemporal model were close to the ones of the longitudinal submodel of the Joint Spatiotemporal model (21 out of 23 credibility intervals overlapped) (Table C.4 in appendix) On the No-link dataset, parameters were also close except for the mixing matrix which might be due to the number of sources overestimated (21 out of 23 credibility intervals overlapped, with 6 disjoint credibility intervals found for the mixing matrix coefficients) (see section 4.4.2.1) (Table C.5 in appendix). This was coherent with the results of section 4.4.2.2 as we observed that the longitudinal results seemed correct even when the shared latent age hypothesis was invalid.

Scale of the Weibull distribution On the Real-like dataset, the values of the Weibull scale corrected by t_0 of the Joint cause-specific Spatiotemporal model were close to the Weibull scale of the cause-specific AFT model for the NIV initiation (5.4 [5.2, 5.7]) and 6.2 [5.7, 6.8]) and the death (5.9 [5.6, 6.2] and 8.4 [7.1, 9.8]) (Table 4.6). On the No-link dataset, the values of the Weibull scale corrected by t_0 of the Joint cause-specific Spatiotemporal model, $(\nu + t_0)$, compared to the one of the Weibull scale of the cause-specific AFT model, were three times bigger for the NIV initiation (14.5 [13.0, 16.8] against 5.8 [5.3, 6.4]) and ten times bigger for the death (92.2 [77.0, 116.6] against 9.1 [7.3, 11.4]) (Table 4.6). Thus really different values of the Weibull scale must be a warning that the shared latent age hypothesis might be invalid.

Shape of the Weibull distribution Depending on the value of the shape of the Weibull distribution (ρ) the hazard function $h(t)$ has different progressions (Jiang and Murthy, 2011):

- $\rho < 1$: indicates that the hazard function decreases over time which happens if the event is more likely to occur at the beginning of the disease,
- $\rho = 1$: indicates that the hazard function is constant over time which might suggest random external events are causing the event,
- $\rho > 1$: indicates that the hazard function increases with time, which happens when the event is more likely to happen as time goes on.

In the Real-like dataset, the trends captured were the same ($\rho > 1$), with values of ρ close between the Joint cause-specific Spatiotemporal model and the cause-specific AFT model for NIV initiation (1.7 [1.6, 1.9] and 2.0 [1.8, 2.3]) and the death (2.9 [2.7, 3.1] and 2.3 [1.9, 2.7]) (Table 4.6).

In the No-link dataset, we observed that the tendencies captured by the two models were not the same: $\rho \leq 1$ for the Joint cause-specific Spatiotemporal model (NIV initiation: 0.9 [0.8, 1.0], death:0.9 [0.9, 1.0]) whereas $\rho > 1$ for the cause-specific AFT model (NIV initiation: 2.4 [2.1, 2.7], death: 2.5 [2.0, 3.0]) (Table 4.6). Here, at least for the death, we expected the hazard function to increase with time ($\rho > 1$).

The fact that the Joint cause-specific Spatiotemporal model did not capture this behaviour in addition to different behaviour compared to the cause-specific AFT model, should be a supplementary warning that the shared latent age hypothesis might be invalid. Indeed, the latent age should not change the hazard function tendency.

To conclude, to test if the shared latent age hypothesis seems valid, we recommend comparing the results of the Joint cause-specific Spatiotemporal model with those of the cause-specific AFT model:

- ν should be close to $\nu + t_0$,
- ρ should lead to the same types of hazard function.

TABLE 4.6: Comparison of the parameters estimated by the Joint cause-specific Spatiotemporal model and the cause-specific AFT model on simulated data

Legend: NIV: Non Invasive Ventilation initiation, Parameters: parameters of the Weibull distribution with ν the scale, ρ the shape, and t_0 the estimated reference time, Simulated: values of the parameter used to simulate the data, Joint Spatiotemporal: Joint cause-specific Spatiotemporal model, Cause-specific AFT: Cause-specific Accelerate Failure Time model. Simulated values ($\nu_l + t_0$, ρ_l) for the Real-like dataset: NIV initiation (4.4, 1.7), death (5.3 2.8) and for the No-link dataset: NIV initiation (4.1, 1.7), death (4.9, 2.8).

		Weibull scale		Weibull shape	
		Spatiotemporal ($\nu_l + t_0$)	Cause-specific AFT(ν_l)	Spatiotemporal (ρ_l)	Cause-specific AFT(ρ_l)
Real-like	NIV	5.4 [5.2, 5.7]	6.2 [5.7, 6.8]	1.7 [1.6, 1.9]	2.0 [1.8, 2.3]
	Death	5.9 [5.6, 6.2]	8.4 [7.1, 9.8]	2.9 [2.7, 3.1]	2.3 [1.9, 2.7]
No-link	NIV	14.5 [13.0, 16.8]	5.8 [5.3, 6.4]	0.9 [0.8, 1.0]	2.4 [2.1, 2.7]
	Death	92.2 [77.0, 116.6]	9.1 [7.3, 11.4]	0.9 [0.9, 1.0]	2.5 [2.0, 3.0]

4.4.3 Intermediate conclusion

The number of sources should be determined using conditional WAIC or other measures of fit. To use the model, a main hypothesis should be made: the longitudinal and the survival processes are linked by a latent disease age. To assess this hypothesis, we recommend comparing the results of the Joint cause-specific model to those of a cause-specific AFT model comparing the results of the Joint cause-specific Spatiotemporal model with those of the cause-specific AFT model:

- ν should be close to $\nu + t_0$,
- ρ should lead to the same types of hazard function.

When the hypothesis of a shared latent disease age seems respected (close values between the Joint cause-specific survival submodel and the cause-specific AFT model), and after choosing the right number of sources, the model well reconstructed the data with 18, out of

32, of the 95% credibility interval of the distribution that contained the simulated parameters and absolute biases lower than 30% otherwise (Table 4.5). The model well reconstructed the latent disease age (temporal random effects) and captured most of the significant association between the ordering of longitudinal outcomes and the occurrence of events (ζ).

4.5 Benchmark

The objective of this section was to evaluate if our model could improve modelling precision compared to the cause-specific JMbayes2.

4.5.1 Method

We made a 10-fold cross-validation (train 90% - test 10%) on the Benchmark dataset. For each Joint cause-specific Spatiotemporal model, we first trained the Longitudinal model for 2,000 iterations. Then, we ran the Joint Spatiotemporal model for 70,000 iterations (on average 7 hours) using the values of the Longitudinal model as initialisation (with a Robbins-Monroe convergence phase on the 10,000 last iterations (Robbins and Monro, 1951)). The cause-specific JMbayes2 model ran for 25,000 iterations (on average 3 hours and a half). The Longitudinal model was also run for 70,000 iterations (with a Robbins-Monroe convergence phase on the 10,000 last iterations (Robbins and Monro, 1951)).

We first compared the fitted models using conditional WAIC. It was output from the cause-specific JMbayes2 model and we computed it for the Joint cause-specific Spatiotemporal model as described in appendix C.1.

Then, we compared them using prediction of both longitudinal and survival outcomes: we personalised the models on the two first visits of the patients from the tests set and made predictions on the remaining. As in section 3.3.2.2, the goodness of longitudinal predictions was assessed using absolute errors for each of the three longitudinal outcomes. We assessed the goodness of survival predictions in ordering events using the C-index at 1 and 1.5 years and the mean cumulative dynamic AUC at 1 and 1.5 years (see section 3.3.2.2). We used the Integrated Brier Score (IBS) to evaluate the precision of predictions of survival predictions (see section 3.3.2.2). All the survival metrics were computed using the Python package sksurv (Pölsterl, 2020). The predictions were compared using a Wilcoxon signed-rank test with a Bonferroni adjustment.

4.5.2 Results

The Joint cause-specific Spatiotemporal model got significantly lower conditional WAIC, across the 10-fold cross-validation, compared to the cause-specific JMbayes2 model: -108,737 (586) against -101,778 (1,228) (p-value = 1.95e-3).

Longitudinal results 12,197 longitudinal predictions were made at 0.55 (0.47) years from the last visit. The Longitudinal model was significantly better than the Joint cause-specific Spatiotemporal model for all the different outcomes, even though the difference was small: the larger being for gross motor scale with 1.424 (1.331) against 1.414 (1.335) (p-value =

6.5e-30) (Table 4.7). The cause-specific JMbayes2 model got significantly lower absolute bias for two outcomes compared to the Joint cause-specific Spatiotemporal model: for bulbar scale (1.166 (1.233) against 1.187 (1.312) (p-value: 3.4e-02)) and gross motor scale (1.365 (1.288) against 1.424 (1.331) (p-value: 3.4e-02)) (Table 4.7).

Survival results The cause-specific JMBayes2 model got significantly better results for the IBS for both NIV initiation (0.124 (0.015) against 0.131 (0.013) (p-value: 7.8e-03)) and death (0.138 (0.021) against 0.142 (0.02) (p-value: 1.3e-06)) (Table 4.8). For AUC and C-index, although the results of the Joint cause-specific Spatiotemporal model were systematically better, none was significant (Table 4.8).

TABLE 4.7: Absolute bias on the longitudinal outcomes for the Joint cause-specific Spatiotemporal and reference models on PRO-ACT data (Benchmark dataset)

Legend: Joint: the Joint cause-specific Spatiotemporal model, Longitudinal: Spatiotemporal longitudinal model, JMbayes2: joint model with shared random effects with cause-specific survival model estimated using JMbayes2.
Results are presented with the mean (SD) over the 10-fold cross-validation. P-values are computed using a Wilcoxon signed-rank test with Bonferroni between the Joint Temporal model and each of the reference models. The absolute bias should be minimised and the best results are in bold. 12,197 longitudinal predictions were made at 0.55 (0.47) years from the last visit.

	Joint	Longitudinal	p-value	JMbayes2	p-value
Bulbar	1.187 (1.312)	1.179 (1.301)	2.8e-18	**1.166 (1.233)**	3.4e-02
Fine motor	1.510 (1.425)	**1.499 (1.417)**	5.1e-24	1.502 (1.397)	9.8e-01
Gross motor	1.424 (1.331)	1.414 (1.335)	6.5e-30	**1.365 (1.288)**	6.1e-08

TABLE 4.8: Event metrics on NIV initiation and death for the Joint cause-specific Spatiotemporal and reference models on PRO-ACT data (Benchmark dataset)

Legend: NIV: Non Invasive Ventilation initiation, Joint: the Joint Temporal model, Longitudinal: Spatiotemporal longitudinal model, JMbayes2: joint model with shared random effects with cause-specific survival model estimated using JMbayes2.
Results are presented with the mean (SD) over the 10-fold cross-validation. P-values are computed using a Wilcoxon signed-rank test with Bonferroni. ↓means that the metric should be minimised and ↑maximised. Results in bold are the best for each metric for each event.

	NIV			Death		
	Joint	JMbayes2	p-value	Joint	JMbayes2	p-value
IBS ↓	0.131 (0.013)	**0.124 (0.015)**	7.8e-03	0.142 (0.02)	**0.138 (0.021)**	1.3e-06
AUC ↑	**0.642 (0.085)**	0.633 (0.091)	1.0e+00	**0.719 (0.101)**	0.695 (0.107)	1.7e-01
C-index 1.0y ↑	**0.654 (0.043)**	0.632 (0.046)	5.9e-02	**0.654 (0.042)**	0.637 (0.042)	1.6e-01
C-index 1.5y ↑	**0.654 (0.044)**	0.642 (0.048)	2.2e-01	**0.655 (0.042)**	0.642 (0.045)	1.2e-01

4.6 Application

The objective of this section was to see how the modelling of NIV initiation could impact the results described in chapter 2.

4.6.1 Method

We assessed the association of longitudinal and survival processes and chose the number of sources with the method described in section 4.4.1. We ran one model on the Analysis PRO-ACT dataset for 70,000 iterations (with a Robbins-Monroe convergence phase on the 10,000 last iterations (Robbins and Monro, 1951)). We extracted from the individual posteriors the mean of the random effects from 10,000 iterations (between 50,000 and 60,000 before the Robin-Monroe scheme (Robbins and Monro, 1951)).

Then, to better characterize the heterogeneity associated with sex (man/woman) and onset site (spinal/bulbar), we studied the distribution of random effects according to four subgroups using ANOVA with Bonferroni correction.

4.6.2 Results

4.6.2.1 Model selection

The number of sources must be inferior or equal to the number of outcomes studied minus one, thus in our case, we tested one and two sources. We computed the conditional WAIC for the model with one (-79,491) and two sources (-116,937) and chose to use two sources. The parameters of the Weibull distribution estimated were close to the one of the competing risk analysis alone (Table 4.9) which enabled us to validate the hypothesis of the shared latent age.

TABLE 4.9: Comparison of the parameters estimated by the Joint cause-specific Spatiotemporal model and the cause specific AFT model on the Analysis dataset

Legend: NIV: Non Invasive Ventilation initiation, Death: Death, Parameters: parameters of the Weibull distribution with ν the scale, ρ the shape and t_0 the estimated reference time, Simulated: values of the parameter used to simulate the data, Joint Spatiotemporal: Joint cause-specific Spatiotemporal model, Cause-specific AFT: Cause-specific Accelerate Failure Time model

Event	Weibull scale		Weibull shape	
	Spatiotemporal ($\nu_l + t_0$)	Cause-specific AFT(ν_l)	Spatiotemporal (ρ_l)	Cause-specific AFT(ρ_l)
NIV	4.6 [4.5, 4.8]	4.9 [4.6, 5.1]	1.9 [1.8, 2.0]	2.1 [1.9, 2.2]
Death	5.1 [5.0, 5.3]	6.6 [6.1, 7.2]	3.4 [3.2, 3.6]	2.3 [2.1, 2.6]

4.6.2.2 Estimation accuracy analysis

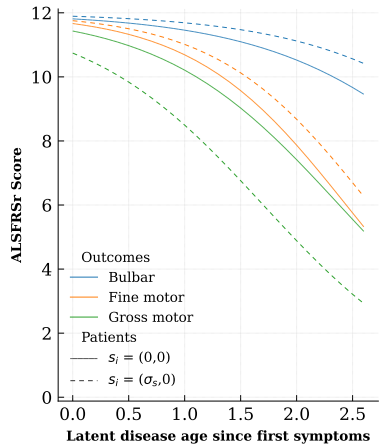
Here we mainly focus on the fixed parameters that enable linking the longitudinal and the survival processes, as it is the main novelty of the Joint cause-specific model, but all the fixed effects of the model are available in Table C.7 in appendix.

The individual spatial variability, is captured by the sources s_i which impact the disease presentation through the mixing matrix A (dimension reduction) and the occurrence of events through the hazard ratio coefficients ζ (proportional impact of the sources on the hazard). Thus studying the dimensions of the sources can give insight into the distribution of the variability. Note that to be easier to interpret coefficients of the mixing matrix are corrected by the speed (v_0) to be on a time unit.

An increase of one standard deviation on the first dimension of the disease presentation (first source $s_{(i,1)} + \sigma_s$) (Figure 4.1 Panel A) corresponded to a higher risk of NIV (HR: 1.2 [1.1, 1.3]), a higher risk of death or tracheostomy (HR: 1.4 [1.2, 1.5]), a later bulbar progression (delay: 6. 9 [6.0, 7.8] months), a later fine motor progression (delay: 2.4 [2.1, 2.6] months) and an earlier gross motor progression (delay: -8.3 [-8.7, -8.0] months). An increase of one standard deviation on the second dimension (second source $s_{(i,2)}+\sigma_s$) (Figure 4.1 Panel B) corresponded to a lower risk of NIV (HR: 0.9 [0.9, 1.0]), a lower risk of death or tracheostomy (HR: 0.8 [0.7, 0.9]), a later bulbar progression (delay: 16.7 [16.0, 17.4] months), earlier fine motor progression (delay: -6.5 [-6.9, -6.2] months) and an earlier gross motor progression (delay: -3.9 [-4.3, -3.5] months).

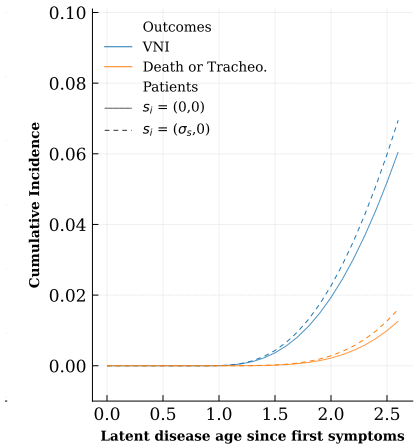
A ***Influence of disease presentation first dimension***

Longitudinal Outcomes



Outcomes	Delay to mean ($A_{k,0}v_{0,k}^{-1}$) *(months)*
Bulbar	6. 9 [6.0, 7.8]
Fine Motor	2.4 [2.1, 2.6]
Gross motor	-8.3 [-8.7, -8.0]

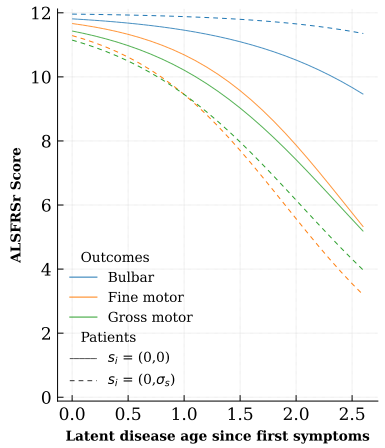
Events Cumulative Incidence



Outcomes	Hazard ratio ($\zeta_{l,0}$)
VNI	1.2 [1.1, 1.3]
Death or Tracheo.	1.4 [1.2, 1.5]

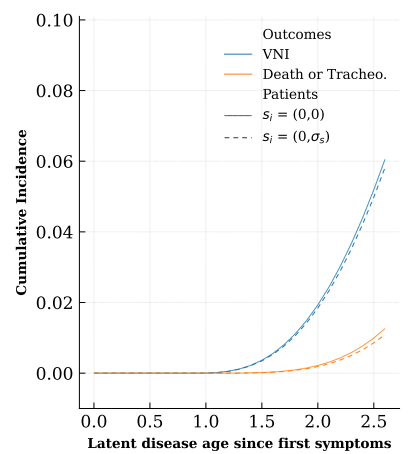
B ***Influence of disease presentation second dimension***

Longitudinal Outcomes



Outcomes	Delay to mean ($A_{k,1}v_{0,k}^{-1}$) *(months)*
Bulbar	16.7 [16.0, 17.4]
Fine Motor	-6.5 [-6.9, -6.2]
Gross motor	-3.9 [-4.3, -3.5]

Events Cumulative Incidence



Outcomes	*Hazard ratio ($\zeta_{l,1}$)*
VNI	0.9 [0.9, 1.0]
Death or Tracheo.	0.8 [0.7, 0.9]

FIGURE 4.1: Influence on the progression of the two dimensions of the disease presentation (spatial aspect)

Legend: Graphs represent the population mean random effects for the speed and the time-shift ($\xi_i = 0, \tau_i = t_0$) and different values for the source s_i depending on the latent disease age since first symptoms (interval correspond to the 95% credibility interval of estimated latent disease age across patients). On each panel the mean value for the disease presentation random effect, $s_i = (0, 0)$ (plain line) is compared to a shift of one standard deviation ($\sigma_s = 1$) of the source (dashed line): Panel A on the first dimension and Panel B on the second dimension. The mixing matrix coefficients ($A_{k,m}$) are corrected by the speed ($v_{0,k}$) to be expressed in time units (months). $\zeta_{l,m}$ are the proportional hazard ratios associated with the sources. The influence of the dimension on the disease presentation has been reported with a credibility interval 95%. Longitudinal outcomes: the shifts in the month for each outcome compared to the trajectory with $s_i = (0, 0)$ are presented (a negative value means that the degradation starts earlier). Events cumulative incidence: the hazard ratios are reported (a value above one means a lower survival and a higher risk).

4.6.2.3 Individual temporal variability

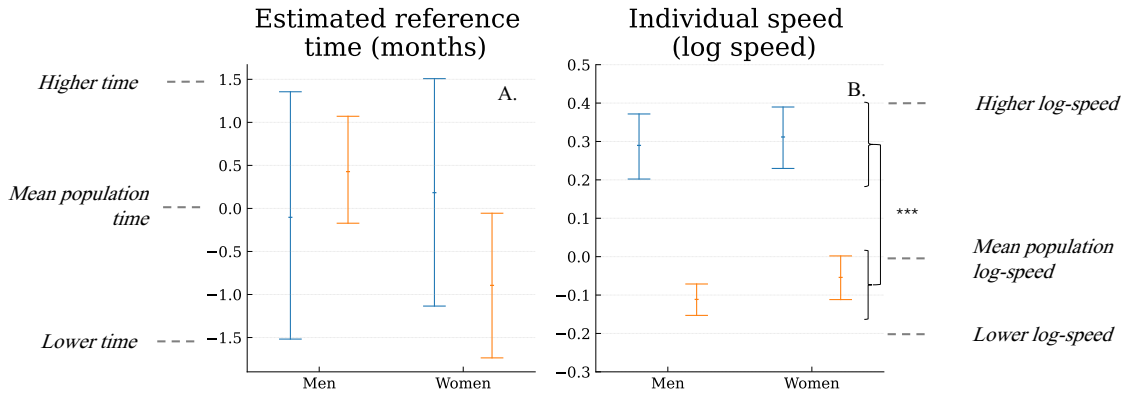


FIGURE 4.2: Individual estimated reference time and speed depending on sex and symptom onset

Legend: Graphs present the mean of random effects distribution for the four subgroups defined by sex (in abscissa men, women) and symptom onset (blue: Bulbar, orange: Spinal) with its confidence interval 95%. Panel A: the vertical axis presents the estimated reference time in months compared to the mean estimated reference time of the whole population. Panel B: The vertical axis presents the log speed compared to the mean log speed of the whole population. ANOVA interaction p-value with Bonferroni correction: (A) 1. estimated reference time, (B) 1. individual log-speed.

Estimated age at disease onset The estimated reference time was not significantly different between the four studied groups (Figure 4.2 A).

Speed of progression We did not find any significant interaction between the onset site and sex for the speed factor of progression (p-value = 1.) (Figure 4.2 B). However, patients with bulbar onset were found to progress 1.47 times faster (95% CI = [1.37, 1.57]) than patients with spinal onset independently of sex.

4.6.2.4 Individual spatial variability

Sources characterise the dimensions of spatial variability, i.e. the order of evolution of longitudinal outcomes and their impact on events. Nevertheless, to describe individual variability on one longitudinal outcome k or event l, space shift ($w_{i,k} = A_k s_i$) and survival shits ($u_{i,l} = \zeta_l s_i$) are usually easier to interpret compared to sources s_i, as they encapsulate the total effect of the spatial variability (see section 4.2.2.1). Note that survival shifts have a proportional impact on the hazard and their interpretation is close to the one of the hazard ratio and will be referred to as the Proportional effect of survival shits on the Hazard (PH). The space shifts were corrected by the speed (v_0) to be on a time unit.

Motor decline ($w_i v_0^{-1}$) We found an interaction between sex and onset site for gross motor (p = 0.034) but not for fine motor scales (p-value = 1.), once corrected for the estimated reference timing and speed of progression (Figure 4.3 A and B).

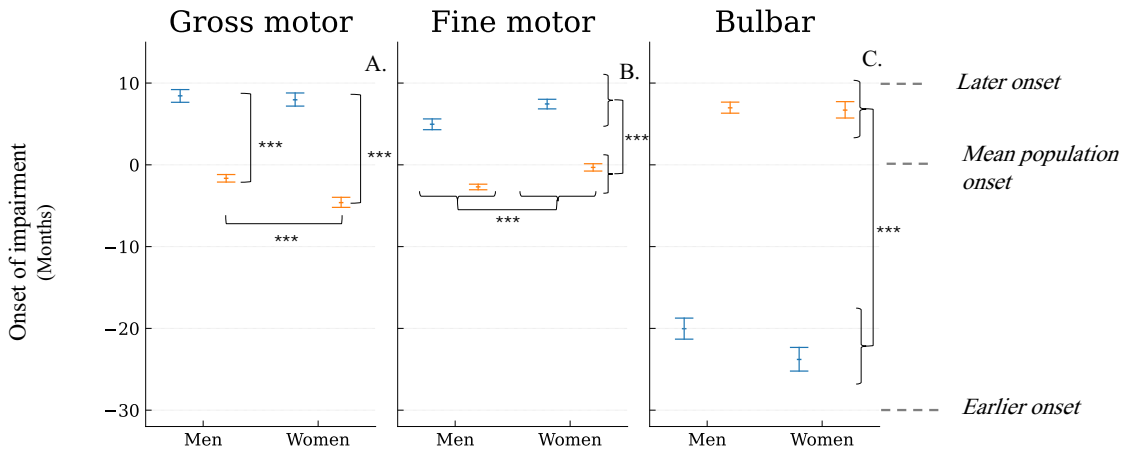


FIGURE 4.3: Individual spatial variability on Longitudinal outcomes

Legend: Graphs present the mean of random effects distribution for the four subgroups defined by sex (in abscissa men, women) and symptom onset (blue: Bulbar, orange: Spinal) with its confidence interval 95%. The vertical axis presents the delay of outcome impairment onset in months compared to the mean onset of the whole population. ANOVA interaction p-value with Bonferroni correction: (A) 0.034 gross motor scale, (B) 1. fine motor scale, (C) 0.15 bulbar scale.

Among patients with spinal onset, ALSFRSr gross motor scale deteriorated 3.0 months later (95% CI = [2.0, 4.0]) in women than in men. However, ALSFRSr fine motor scale changed 2.4 months earlier (95% CI = [1.9, 2.9]) in women than in men, independently of the onset site.

Among men, ALSFRSr gross motor scale changed 10.1 months earlier (95% CI = [8.8, 11.4]) for patients with spinal onset compared to bulbar onset. Among women, the difference decreased to 7.7 months earlier (95% CI = [7.2, 8.3]). ALSFRSr fine motor scale changed 10.2 months earlier (95% CI = [9.0, 11.3]) for patients with spinal onset compared to bulbar onset, independently of the sex.

Bulbar signs decline ($w_i v_0^{-1}$) We did not observe any interaction between sex and onset site for ALSFRSr bulbar scale (p-value =0.15), once corrected for the estimated reference timing and speed of progression (Figure 2.7 C).

ALSFRSr bulbar scale changed 28.6 months later (95% CI = [27.4, 29.8]) for patients with spinal onset compared to bulbar onset, independently of the sex.

Non-Invasive Ventilation Initiation (u_i) After correction for speed and onset, women had a significantly higher risk of NIV initiation compared to men (PH: 1.09 [1.08, 1.11]) (Figure 4.4 A).

Death (u_i) Again, after correction for speed and onset, women had a significantly higher risk of death compared to men (PH: 1.16 [1.14, 1.20]) (Figure 4.4 B).

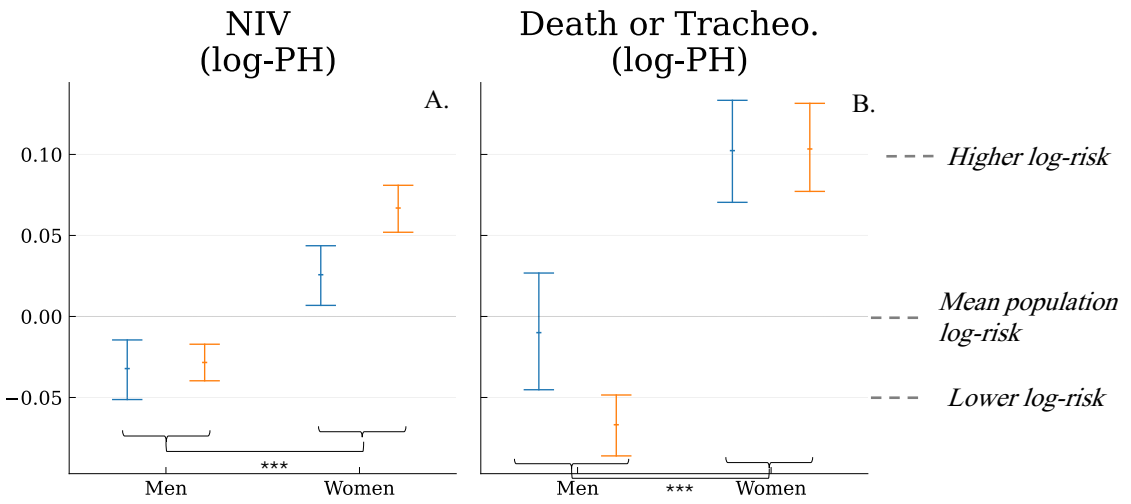


FIGURE 4.4: Individual spatial variability on NIV initiation and death
Legend: Graphs present the mean of random effects distribution for the four subgroups defined by sex (in abscissa men, women) and symptom onset (blue: Bulbar, orange: Spinal) with its confidence interval 95%. The vertical axis presents the log Proportional effect of survival shits on the Hazard compared to the mean of the whole population. ANOVA interaction p-value with Bonferroni correction: (A) 1. NIV initiation log-PH, (B) 1. death

4.7 Discussion

4.7.1 Results interpretation

4.7.1.1 Methodological

We designed the first data-driven multivariate joint cause-specific model. To do so, we used the Longitudinal Spatiotemporal model as longitudinal submodel. The proposed Joint cause-specific Spatiotemporal model realigns both survival and longitudinal observations on a latent disease age (temporal aspect). In addition, it captures the impact of the order of the longitudinal outcome on the survival processes (spatial aspect). This enabled us to overcome the limitation of the joint shared random effects that model the impact of the longitudinal outcomes on survival.

To assess the validity of the shared latent age hypothesis, we recommend comparing the results of the Joint cause-specific model to those of a cause-specific AFT model: the shape of the Weibull distribution should lead to the same types of hazard function and really different values of the scale should be a warning. We advised using the conditional WAIC to find the right number of sources for the model which is the only hyperparameter that the user has to fine tune. When the hypothesis of a shared latent disease age seems respected, and after choosing the right number of sources, the model gave correct results on real-like simulated data with high censure level (20% of observed VNI initiation and 9.1 % of death): with 18, out of 32, of the 95% credibility interval of the distribution that contained the simulated parameters and absolute biases lower than 30% otherwise.

For the benchmark against the reference model, the cause-specific JMBayes2, the Joint cause-specific Spatiotemporal model got a significantly better conditional WAIC on the Benchmark dataset. This could be linked to the fact that the Joint cause-specific Spatiotemporal model has only one individual speed shared for all the outcomes, whereas the cause-specific JMbayes2 model has one per outcome linked with a correlation matrix.

The prediction results were more mitigated. The cause-specific JMBayes2 model got better results for all the longitudinal outcomes. This might be the drawback of one shared individual speed (ξ_i) of the Spatiotemporal model. Nevertheless, this shared individual speed should facilitate application to a higher number of longitudinal outcomes, which is currently a limitation of the cause-specific JMBayes2 model (Devaux et al., 2023; Hickey et al., 2016). Compared to the Joint cause-specific Spatiotemporal model, the Longitudinal Spatiotemporal model got significantly better results. It was already the case with the Joint Temporal model and could be because the longitudinal model is less constrained.

As for the Joint Temporal model with the univ-JMbayes2 model in chapter 3, the cause-specific JMBayes2 model got better results for the event distance metric (IBS). The survival function of the cause-specific JMbayes2 model may exhibit this difference due to its enhanced flexibility, utilizing splines instead of a Weibull function. Even though the Joint cause-specific Spatiotemporal model got systematically better event ordering metrics (C-index and AUC) compared to the cause-specific JMbayes2 model, none was significant. Results were stronger for the Joint Temporal model compared to the uni-JMbayes2 model, for which the event prediction task was a bit easier with only one event to model and 588 (23%) event observations.

4.7.1.2 Clinical

For the clinical application, although the population of this study was younger (due to Mitos selection) and the outcomes did not include respiratory and weight decline proxies as in chapter 2, results on individual speed, gross motor, bulbar and fine motor scales were quite similar. Note that even though results on the functional decline were not significant in our study the same tendency as the one from (Grassano et al., 2024) emerged: women have a faster functional decline. The estimated age at onset could not be compared due to the re-indexation by the time since the onset of symptoms, which seems to have absorbed most of the variability.

Once corrected for speed of functional decline, men had a slower NIV initiation and a slower death or tracheostomy, which did not replicate results from the literature (Grassano et al., 2024; Dibling et al., 2024). This might be linked to the fact that we did not take into account the age at disease onset as a covariate in the survival analysis. Indeed, men are known to start the disease earlier than women (Nona et al., 2023), and a later disease tends to have a worse prognosis (Chio et al., 2009).

4.7.2 Strengths & Weaknesses

4.7.2.1 Methodological

The designed model showed great potential to model a shared disease speed among both longitudinal and survival processes, which could be a nice outcome for causal inference in the future. It must be benchmarked in the context of higher dimensions (Nguyen et al., 2023; Devaux et al., 2023; Hickey et al., 2016).

Nevertheless, its structure makes it harder to model complex associations, such as the impact of the integral of some outcomes, which is possible with JMbayes2. Covariates were also not included in the model, but recent work (Fournier and Durrleman, 2023) paves the way for their integration. Marginal WAIC would have been more robust to compare the models (Millar, 2018), but was computationally consuming due to the current version of the software.

4.7.2.2 Clinical

A strength of our study is that our model was based on a large, multicentric, and longitudinal database. However, the PRO-ACT database may not represent the real-life ALS patient population. Since the database aggregates clinical trial data, there is an over-representation of men, young subjects, patients with spinal onset, and slow progressors. An improvement in dealing with survival data could be done by handling interval and left censored (42 out of 2,177 patients) NIV initiation to reduce potential introduced biases (Leffondré et al., 2013). The main limitation of our study was that we did not include the age at disease onset as a covariate for both NIV initiation and death or tracheostomy, which may be the main reason why we could not replicate the results of (Grassano et al., 2024). Recent improvements of the Spatiotemporal model have enabled taking into account covariates, and such functionality should be added to the joint Spatiotemporal model (Fournier and Durrleman, 2023).

4.8 Conclusion & perspectives

The Joint cause-specific Spatiotemporal model has great potential but must be further tested on high dimensional data. More extensions with covariates would also be needed to better characterise NIV initiation.

Chapter 5

Software development: contribution to reproducibility and reuse

This chapter describes the two main software tools developed during this thesis to obtain the results presented in the previous chapters and to ease future reuse and reproducibility of the work done. First, we implemented the joint models described in the past chapters in an existing open-source software named Leaspy for LEArning Spatiotemporal patterns in PYthon.To do so, a large refactorisation was conducted to ease maintenance and model implementation. Second, we integrated all the data processing into an internal software. All this was a team effort in which I participated.

Contents

5.1 Introduction

Sharing the data and the code used during experiments enables to improve reproductibility and fosters an easier reuse of the work done in other contexts. During this thesis, two contributions were made in that sense.

First, following the call of the literature review of Hickey et al., 2016 and the efforts of the joint modelling community (Proust-Lima, Philipps, and Liquet, 2017; Rizopoulos, Papageorgiou, and Miranda Afonso, 2024; Rustand et al., 2024; Lavalley-Morelle et al., 2024), we implemented the joint models described in the previous chapters in an existing open-source software named Leaspy for LEArning Spatiotemporal patterns in PYthon. Second, getting inspiration from what was done by Routier et al., 2021, we integrated the data processing required for this thesis into a team software to facilitate its future use and improvement.

5.2 Leaspy: an open-source software for disease progression modelling

5.2.1 Context

A software already exists for the longitudinal Spatiotemporal model: Leaspy (`https://gitlab.com/icm-institute/aramislab/leaspy`) and was used in several clinical studies (Koval et al., 2021; Sauty and Durrleman, 2023; Maheux et al., 2023; Landau et al., 2021; Moulaire et al., 2023; Koval et al., 2022; Di Folco et al., 2023). Thus, the objective was to integrate the developed joint models into the existing software.

Doing so, quite a lot of problems were raised. First, base classes (root classes from which others inherited) did not define a proper interface and they contained specific logic, which led to inheritance issues for the introduction of new models. Second, there was a tight coupling between different classes, which implied changing a lot of files across the code when introducing a new model. Third, model specifications were far from the mathematical representation, which again complexified the implementation of new models.

For all those reasons and because there was an urge to introduce several models in addition to the joint models described (Fournier and Durrleman, 2023), a large team effort was started to refactor the existing software and facilitate future model implementation. This section presents the work done, taking as an example the implementation of the joint temporal model (see section 3.2.2).

5.2.2 Refactorisation of the class structure

5.2.2.1 The intuition behind the new structure

Going back to the parameters structure of the different models implemented (longitudinal multivariate: section 2.2.4.1, joint temporal: section 3.2.3.1, joint competing multivariate: section 4.2.3.1), we can see that parameters have different properties. For instance: latent parameters are sampled, model parameters maximise the likelihood, some are shared across

patients, others are individual. Nevertheless, they can all be seen as belonging to a common framework: they are quantities that are computed from other variables or that allow computing other variables (often both). All the Spatiotemporal models, as Bayesian models, are therefore just a set of variables along with a set of dependency relationships between these variables. Every intermediate computation of the model, as the latent age $\psi_i(t)$ (see section 2.2.3.2) or the longitudinal process $\gamma_i(t)$ (see section 2.2.3.3), also corresponds to instantiating new variables that depend on the elements of the computation.

For models to be computationally sound, we needed to have a flow of dependency: a variable cannot depend on a variable that itself depends on it. Said differently, every model should be laid out as a Directed Acyclic Graph (DAG). The graph specifies the properties of each variable, as well as the dependency flow and the update rules. Such structure can optimise computations and has the advantage of being close to the mathematical formulation. In the case of the joint temporal model, the DAG is described in Figure 5.1. This structure was inspired by Bayesian software such as PymC (Oriol et al., 2023).

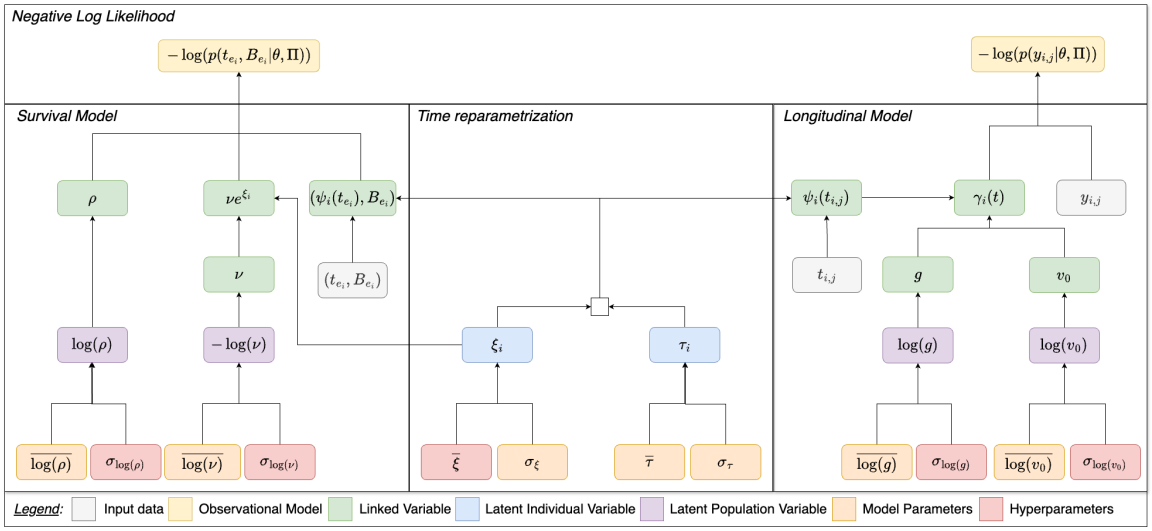


FIGURE 5.1: Directed Acyclic Graph of the joint Temporal model
Legend: individual parameters are indexed by i. θ: model parameters, Π: hyperparameters, ρ: shape of the Weibull distribution, ν: scale of the Weibull distribution, (t_{e_i}, B_{e_i}): time at event and associated boolean to code if it is observed or not, ξ_i: individual speed factor, τ_i: individual estimated reference time, $t_{i,j}$: time at visit j, g: Curve value at t_0: $\frac{1}{1+g}$, v_0: Speed of the logistic curve, $y_{i,j}$: outcome at the visit j. Note that $t_0 = \bar{\tau}$.

In practice, computation did already follow this flow of dependency after each update of variables. But it was in a very "ad-hoc" and model-specific way following long condition blocks spread in multiple places in the code.

5.2.2.2 The new structure

The new structure can be summarised by the diagrams displayed in Figure 5.2 and is further described by the four parts below.

Variables Following the idea that model update follows a DAG, such a *VariableDAG* object was created (Figure 5.2 Variables part). It is a formal object that defines the model through formal variables and their formal relationships.

The different variables stored in the DAG are of multiple types depending on the parameters they represent:

- *DataVariable*: observed data (ex: y, t_e, B_e),
- *Hyperparameter*: parameters set by the user (ex: σ_g),
- *ModelParameter*: parameters that are maximised through the likelihood (ex: t_0),
- *IndividualLatentVariable*: latent individual variables (random effects) that are sampled (ex: ξ_i),
- *PopulationLatentVariable*: latent population variables (fixed effects) that are sampled (ex: g, v_0),
- *LinkedVariable*: variables obtained only from the computation of other variables (ex: $\psi_i(t)$)),

Each model is defined by a *State* object that contains a DAG and the current values of each variable. It can be thought about as a dictionary, with set and get methods to handle the propagation of values defined by the DAG. This object has fork and clone methods to facilitate Gibbs acceptation scheme (Casella and George, 1992) during the calibration (see section 2.2.5 for more details).

Observation models We defined a class named *ObservationModel* that given the model and some observed data will provide the linked variables regarding observations and their attachment to the model (the negative log-likelihood - nll - to be minimized) associated with a well-defined distribution.

Distribution Latent variables and observation models are based on a probability distribution that is defined by a specific class, all inherited from the same structure.

Models The model class (that inherited from *BaseModel*)) link the state, the observation model, and the metric (*LogMixing* and *LinMixing*) that defines the shape of the curves as defined in section 2.2.3.3.

5.2.2.3 Non-regression evaluation

In the initial version of the package, quite a lot of tests were already implemented: unit tests which enable to check that each method of each class does what it is supposed to do, and functional tests which verify that the implemented models return the right output after the different estimations. All the tests are run after each change in the package to make sure nothing was broken accidentally. Nevertheless, these tests were done on a few iterations and did not wait for convergence, due to time considerations. The refactoring

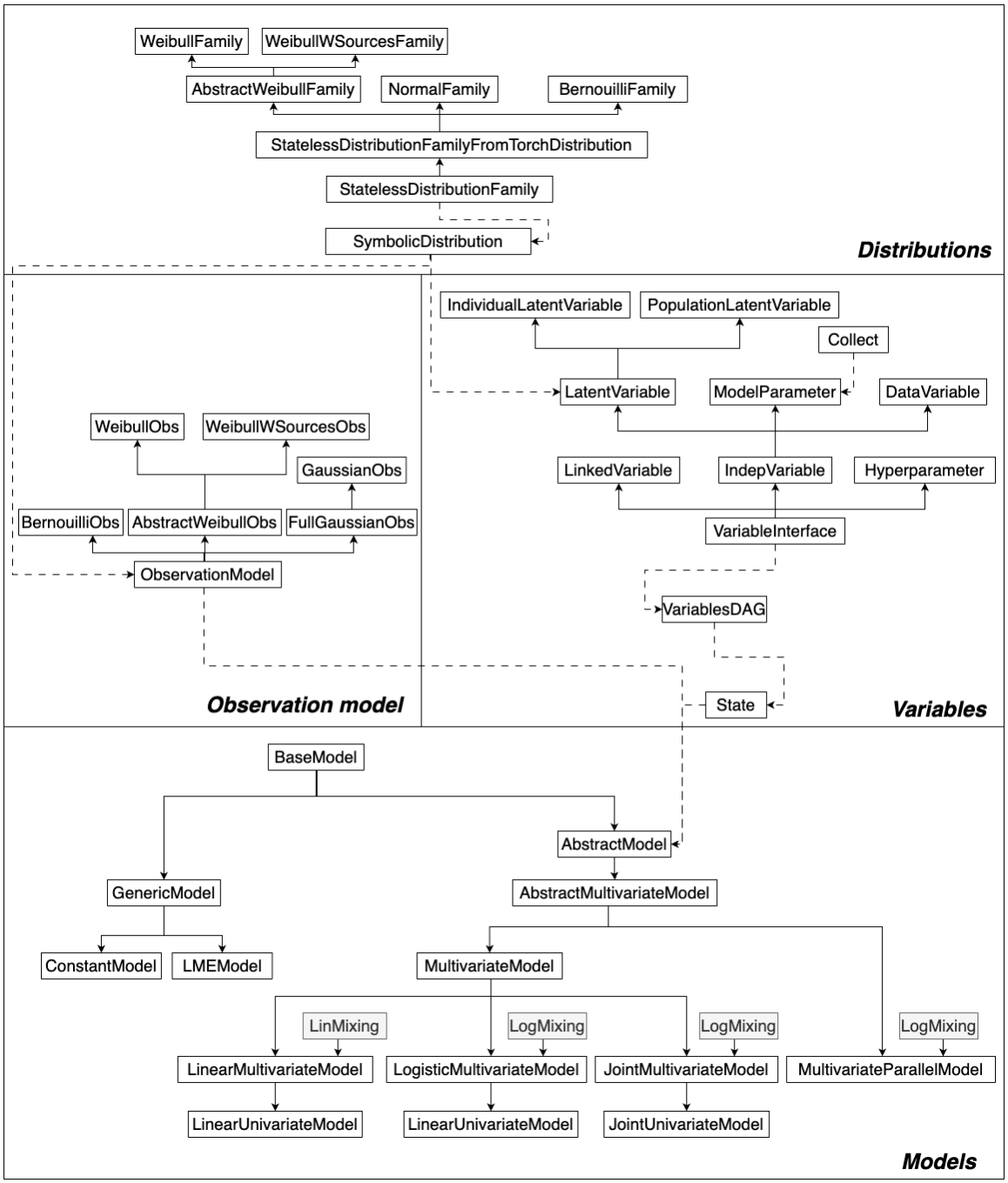


FIGURE 5.2: Simplified Class Diagram structure of Leaspy software

Legend: Dashed: the class (arrow start) is an attribute of the other class (arrow end), Plain: the class (arrow end) inherited from the parent class (arrow (inherited from the method of the parent class and can have its own), WeibullFamily: Weibull distribution from latent age for right censored data, WeibullWSourcesFamily: Weibull distribution from latent age for right censored data with a proportional effect of sources on the hazard, WeibullObs: Observational model with WeibullFamily distribution, WeibullWSourcesObs: Observational model with WeibullWSourcesFamily distribution, Collect: class that contains the update rules for the model parameters (sufficient statistics section 2.2.5), LinMixing/LogMixing: these classes enable to define the shape of the curve linear or logistic, LMEModel: Linear mixed-effect model

of the sampling changed the order of the latent variables, which broke the tests. Thus, to

enable the refactored version to pass the tests, we had to hardcode the order of the variables sampled. Tests are meant to be updated at the end of the refactorisation to remove the hardcode order of the variables sampled.

We have also reproduced different past experiments to make sure that the models converged at the same place.

5.2.3 A simplified implementation: the example of the joint temporal model

Thanks to the refactorisation adding a new model can be summarised by the following steps:

1. **Update input data structure:** Leaspy has a specific class to handle data. Thus, if the new model handles data that were not handled before, the data class must be updated. This was the case for the joint temporal model, as before only longitudinal data were allowed. Changes were made to enable survival data in the class structure.

2. **Create an observation model:** Then if necessary a new observation model can be created to define the link between the observed data and the new model. In the case of the joint temporal model, a new observation model following a Weibull distribution for right-censored survival data was created.

3. **Specify the DAG:** A new model class can then be created and its DAG should be specified. For the joint temporal model, this consisted of defining the DAG described in Figure 5.1.

4. **Write some tests:** When the model has been validated on simulated data, some functional tests must be added to avoid future involuntary changes.

The joint temporal model was coded in both versions (pre and post-refactorisation) which enabled us to measure the improvement for future users. The merge request in the first version was spread over a large number of files whereas the second one was much more localised and mainly followed the above steps. Both merge requests are available online at `https://gitlab.com/icm-institute/aramislab/leaspy/-/merge_requests/104/diffs` and `https://gitlab.com/icm-institute/aramislab/leaspy/-/merge_requests/114/diffs`.

5.3 Data converters: extracting and cleaning data

5.3.1 Context

To ease data processing, conventions exist on how to store longitudinal and survival datasets. For longitudinal datasets, two formats exist: the long format for which each row is one visit of a given patient and columns the outcomes, and the wide format for which each row is a patient and columns represent the value of the outcome at each visit. The long format is by far the most used in clinical datasets. For survival data, also several formats exist (Dugué et al., 2016). For events left or right censored, a time is stored jointly with a boolean. For interval-censored events, the two bounds times of the interval are stored.

With these conventions, the value used for visit indexing can vary, the units of measurement may not be consistent and some artefacts may exist in the databases. Thus, each person tends to create its own cleaning pipeline which may complexify reproducibility.

Doing so, and getting inspiration from Routier et al., 2021, we defined a stricter format and created converters for public datasets. The software aims to enable collaborators to spend less time on data processing, facilitate data sharing, and improve the reproducibility of experiments.

5.3.2 From raw to clean data

5.3.2.1 Dataset format

There are different ways to index the visits: by their number, by their time to the baseline, by the age of the patient ... In some studies visits are not equally spaced thus indexing with the number of the visit does not make sense, a time would be more reliable. Nevertheless, we have seen that symptom onset is not a reliable time reference, and age at baseline may not be representative of the same stage of the disease for all patients. Thus, we chose to index the visit thanks to the age of the patient stored as a float. In addition to time-varying outcomes, patients can be characterised by fixed covariates that do not vary across visits, such as sex. To store them in addition to longitudinal outcomes in the same dataset, they are repeated at each visit but must be equal for one patient.

Like fixed covariates, events were repeated at each visit of the patient. Depending on the extraction method for events (through longitudinal scores or event database) they have different censoring types. Thus, they were stored in different formats (interval censored or right censored).

5.3.2.2 Conventions

To perform automatic tests at the end of data cleaning, each new variable cleaned (outcome, covariate) is added to a convention repertory. It includes a range (minimum to maximum values often defined as worst to best) as well as a type of outcome that can be an item (*CategoricalItem*), clinical score (*NumericCompoundItem*), or a biomarker (*NumericItem*). The class diagram is described in Figure 5.3.

5.3.2.3 Converters

The idea was to create a structure that automatically processes the data so that the code could be shared and run on a machine without directly sharing the data. An abstract class of converter was created and then each converter specific to a dataset can inherited from it. Each converter overloads *get_clean_dataset_impl* and returns, after a batch of automatic tests, an object containing the dataset. Eight converters were created to process datasets from different diseases: ALS, Alzheimer's disease, Huntington, and Parkinson's disease. They are all summarised in Figure 5.3.

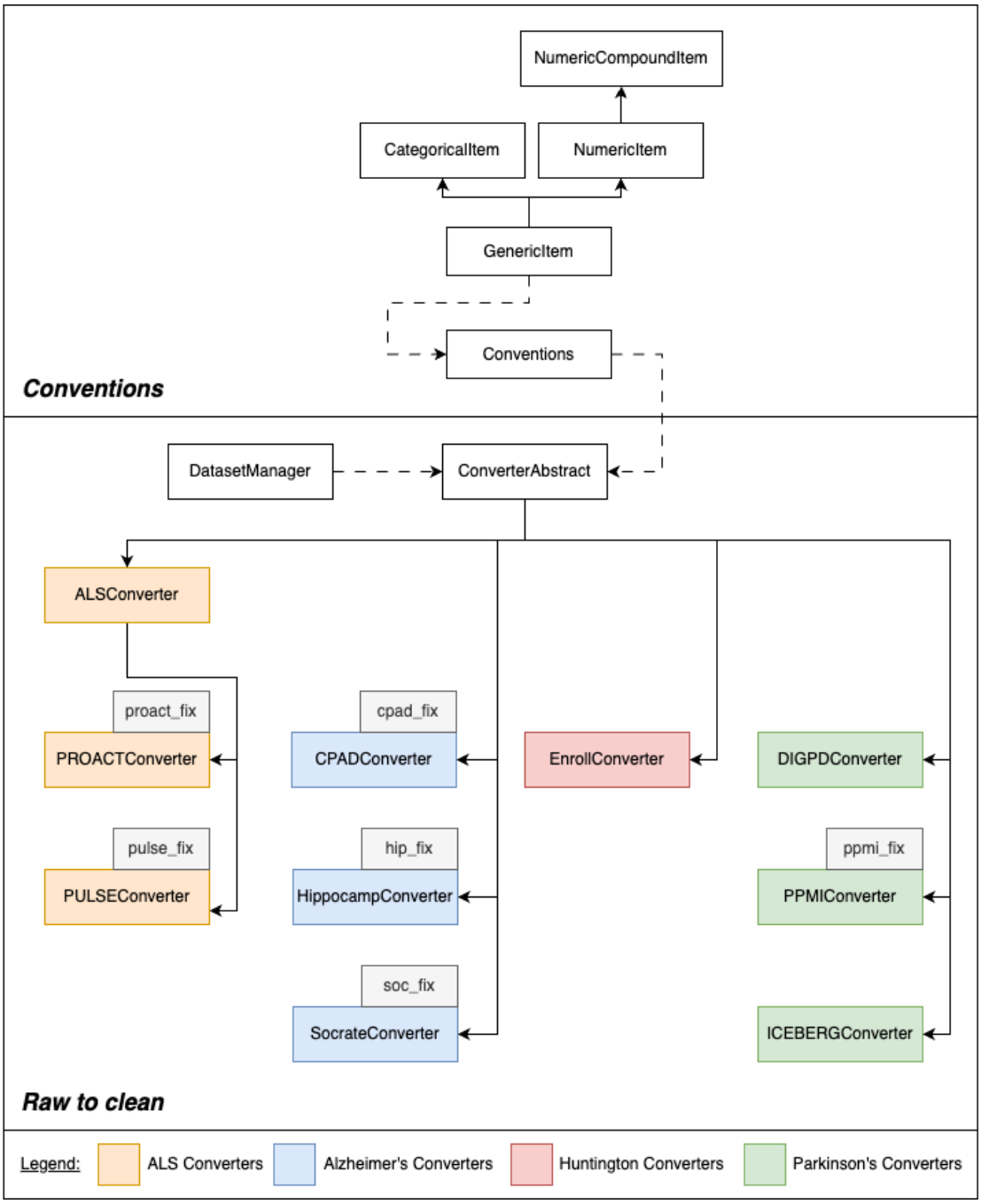


FIGURE 5.3: Simplified Class Diagram structure of the converter software
Legend: Dashed: the object is an attribute, Plain: inherited from the class, xx_fix: json files that contain some individual fix for patients that have strange values.

5.3.3 The example of PRO-ACT converter

The PRO-ACT database is composed of 16 files in long format, with visits indexed by a relative timing to the baseline of the file. In addition, columns are not typed, and some artefacts can be found.

Different outcomes were directly collected from the database (the items of the ALSFRS(r) (Rooney et al., 2017), in opposition to computed ones with for instance a simple sum

(ALSFSR total or Mitos (Fang et al., 2017)). Nevertheless, some values available were already composite scores and their method of computation was not specified in the database, which could lead to centre biases if it was computed at the center level in further analysis. It was for example the case with the normalised Forced Vital Capacity (FVC), for which several normalisation techniques exist (Degens and Merget, 2008). We recomputed with a homogenous formula the normalised FVC at each visit (Degens and Merget, 2008; Brändli et al., 1996). The difference with the version extracted from the database was of 8.2 (5.7) %. Due to the multicentric setting of the PRO-ACT database, weight was collected in different units. Nevertheless, some weights had no units associated, or only for a subset of the visits. Differences between the values in each unit were not large enough to impute data without labels. Thus we had to put aside weights of patients with no labels or inconsistent labels across visits. All the different visits for the different outcomes came from different files and were merged. After merging, we ended up with visits indexed by different ages, sometimes closer than a day. In such cases, we merged them as one unique visit, if the values of the outcomes allow it.

We considered that height could be seen as a covariate for patients older than 25 years, which represent the large majority of the patients in the dataset. As for the weight, we faced an issue with units. There is a factor of 2.54 to pass from inches to centimetres. Thus only the patients with a height between 100 and 110 cm could be difficult to label (100 to 110 cm or 254 or 279 cm). Hopefully, there were no such patients. Then we needed to homogenise the height over the visits for each patient as they were patients with heights that vary up to 20 cm. Sex labelling was consistent. For the ALS onset type, we grouped the different available labels: *LIMB* and *SPINE* for spinal patients and *BULBAR* for bulbar patients. If both were present the values were concatenated.

We also extracted events from the database. Extraction of the death was quite simple as all the deaths were stored in one file. For patients without observed death, we censored their deaths with the date of the last visit. Extraction of Non-Invasive Ventilation (NIV), gastrostomy and tracheostomy was a bit more complex. The use of NIV is coded in item 12 of ALSFRSr and item 10 of ALSFRS. We were thus able to extract interval-censored data thanks to the value of these items taking the first change into account. The same kind of procedure was followed using item 12 of ALSFRSr for the tracheostomy and items 6a and 6b for the gastrostomy. As tracheostomy and gastrostomy are also medical interventions, some codes were available in one of the tables using the Medical Dictionary for Regulatory Activities (MDRA). After analysis of the different labels, we have decided to keep the Lowest Level Term, displayed in Figures D.1 and D.2 in the appendix, to extract an exact date of gastrostomy and tracheostomy. Thus for gastrostomy and tracheostomy two types of events were extracted for some patients: interval-censored events and right-censored events.

From an implementation point of view the cleaning of the database can be separated into three main phases. First, the extraction of fixed covariates, then the extraction of all the longitudinal outcomes and finally the extraction of events.

5.4 Personal contribution

Since the beginning of my PhD, I have been involved in the Leaspy project and participated in its development. Over the past three years, I have made many contributions, both in terms of software development and project management.

Most of the implementation of the refactored existing code was done by Nicolas Gensollen and Etienne Maheux. However, I have actively participated in creating short and midterm roadmaps, reviewing the merge requests, and designing the refactored software. I have also implemented the different joint models in the refactored version.

I participated marginally in the conception of the whole structure of the converter, but I implemented inside the structure all the converters related to ALS.

5.5 Conclusion & perspectives

We contributed to ease the reproducibility and the repurposing of the work done in this thesis.

With the members of my team, we have refactored the existing open-source software to ease maintenance and future model implementation. The main historical models are already available in the package, as well as the new joint models described in this thesis. Nevertheless, some work remains in implementing an automatic diagnosis routine of the calibration estimation, with graphs and key summary statistics including metrics to compare models (WAIC (Watanabe, 2010)). This work may also enable easier implementation of more modern samplers to speed up convergence (Betancourt, 2018; Hoffman and Gelman, 2014).

The data processing pipeline can always be improved by including the processing of more publicly available datasets. Some work could also be done to convert it into an open-source software while maintaining patient privacy.

Conclusion and perspectives

Conclusion

Progression heterogeneity in chronic diseases such as Amyotrophic Lateral Sclerosis (ALS) is a significant obstacle to developing effective treatments. Leveraging the growing wealth of large databases through modelling can help better understand it. However, the data collected only offer access to partial trajectories needing to be realigned to reconstruct a comprehensive disease progression.

This thesis aimed to model the heterogeneity of ALS progression, by capturing both longitudinal and survival processes, while considering the specificity of the available data.

After identifying the need in the literature for a joint data-driven model to handle ALS data, we have extended an existing longitudinal one, the Spatiotemporal model, to fill this gap. The strength of the Spatiotemporal model is its ability to synchronise patients onto a common disease timeline (temporal aspect) thanks to a latent disease age, while also capturing the remaining variability through parameters that account for outcome ordering (spatial aspect).

We first associated the Spatiotemporal model with a survival model that estimates a Weibull survival model from its latent disease age, creating a univariate Joint Temporal model. Our model was validated on simulated data and exhibited on PRO-ACT data significantly better performance in terms of absolute bias and mean cumulative AUC for right-censored events compared to a state-of-the-art shared random effect model.

Then, we coupled the multivariate Spatiotemporal model with a cause-specific Weibull survival model from the latent disease age, incorporating spatial parameters with a Cox proportional effect on the hazard to create a Joint cause-specific Spatiotemporal model. The Joint cause-specific Spatiotemporal model achieved similar performance to the state-of-the-art model while capturing an underlying shared latent process, the latent disease age, whereas the state-of-the-art models the impact of longitudinal outcomes on survival.

Thanks to these models, we contributed to characterise ALS heterogeneity. We applied the Spatiotemporal model to explore how the interaction between sex and onset site (spinal or bulbar) impacts the progression of ALS patients. Using the 1,438 patients selected from the PRO-ACT database, we demonstrated a significant influence of both sex and onset site on six longitudinal outcomes monitoring the functional and respiratory decline in addition to Body Mass Index.

We then complemented this analysis with the newly created Joint cause-specific Spatiotemporal model to analyse the initiation of Non-Invasive Ventilation and death. Our

analysis revealed that women were more at risk than men of NIV initiation and death, after adjusting for individual latent disease age and onset site, but should be further assessed with the correction by the age at first symptoms.

Contributions were made to ease the reuse of the new models developed, by starting a large refactoring of an existing open-source software named Leaspy for LEArning Spatiotemporal patterns in Python and implementing the model developed in this thesis in the new structure. All the data processing needed for this thesis was also integrated into an internal data processing software for publicly available datasets to facilitate reproducibility.

Perspectives

Some work remains to validate and extend the Joint cause-specific Spatiotemporal model. First, further studies to better understand the underestimation of the speed should be conducted including the estimating the influence of the longitudinal noise on the model. Then, a benchmark in the context of chronic disease to see if the Joint Spatiotemporal model could handle high-dimension data in line with the work of Nguyen et al., 2023 should be conducted. Nevertheless, to do so, some efforts must be made to improve the samplers of the model to speed up the convergence (Betancourt, 2018; Hoffman and Gelman, 2014). Then, getting inspiration from (Rouanet et al., 2016; Yang et al., 2023), the Joint cause-specific Spatiotemporal model could be extended to handle interval and left censored data to better analyse NIV initiation or any event that could be extracted from clinical scores for example. Finally, studying the compatibility of the Joint cause-specific Spatiotemporal model with other recent extensions of the Spatiotemporal model that enable modelling ordinal data (Poulet and Durrleman, 2023b), different subgroups through mixture modelling (Poulet and Durrleman, 2023a) or covariates influence (Fournier and Durrleman, 2023) could also be of interest.

Joint model could also be used in clinical trials to improve power through the correction with prognostic scores (Ai, Street, and Francisco, 2021; FDA, 2023) or better-selecting homogenous population (Maheux et al., 2023). In causal inference, joint models may offer an alternative to principal stratification in the context of truncation by death when a longitudinal outcome is the outcome of interest (McNealis, Moodie, and Dean, 2024; Shardell and Ferrucci, 2018; Kurland et al., 2009).

For now, the Spatiotemporal model does not enable modelling changes of dynamics in longitudinal progression. Modelling jointly to longitudinal data, recurrent or multiple events could be a way to introduce them in the model (Kim et al., 2012; Rouanet et al., 2016; Ferrer et al., 2016; Mauguen et al., 2013; Musoro, Geskus, and Zwinderman, 2015; Tong, Liu, and Pena, 2022). Such work could have direct applications, among which, studying the progression changes after attacks in multiple sclerosis or after strokes in CADASIL.

Appendix A

Model equations

In this appendix all the formulas related to the estimation of the models are grouped.

Contents

A.1 Likelihood

As a reminder, note that there are N patients indexed by i and each has n_i visits indexed by j. Note that $t_0 = \overline{\tau}$. Note that the detailed probabilities must be integrated over the latent parameters to get the real likelihood formula.

A.1.1 Longitudinal Spatiotemporal model

k indexed the K outcomes and m the N_s sources (see section 2.2.1), parameters are defined in section 2.2.3.4. Extracted from Koval, 2020 p.175.

$$
\begin{aligned}
\log p(y, z \mid \theta, \Pi) = & \sum_{i,j,k} -\log\left(\sigma_k\sqrt{2\pi}\right) - \frac{1}{2\sigma_k^2}\left(y_{i,j,k} - \gamma_{i,k}\left(t_{i,j,k}\right)\right)^2 \\
- & \sum_{k} \log\left(\sigma_{\tilde{g}}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{g}}^2}\left(\tilde{g}_k - \overline{\tilde{g}}_k\right)^2 \\
- & \sum_{k} \log\left(\sigma_{\tilde{v}_0}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{v}_0}^2}\left(\tilde{v}_{0,k} - \overline{\tilde{v}}_{0,k}\right)^2 \\
- & (K-1)N_s \log(\sigma_\beta\sqrt{2\pi}) - \frac{1}{2\sigma_\beta^2}\sum_{o,m}(\beta_{o,m} - \overline{\beta}_{o,m}) \\
- & N \log\left(\sigma_\tau\sqrt{2\pi}\right) - \frac{1}{2\sigma_\tau^2}\sum_{i}(\tau_i - \overline{\tau})^2 \\
- & N \log\left(\sigma_\xi\sqrt{2\pi}\right) - \frac{1}{2\sigma_\xi^2}\sum_{i}(\xi_i - \overline{\xi})^2 \\
- & NN_s \log\left(\sigma_s\sqrt{2\pi}\right) - \frac{1}{2\sigma_s^2}\sum_{i}\sum_{m}^{N_s}(s_{i,m} - \overline{s})^2
\end{aligned}
$$

A.1.2 Joint Temporal model

T_e is the time of observation of the event and B_e is the associated boolean whether the event was observed or not (see section 3.2.1). Parameters are defined in section 3.2.3.1.

$$
\begin{aligned}
\log p((y, T_e, B_e), z \mid \theta, \Pi) = & \sum_{i,j} -\log\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2\sigma^2}\left(y_{i,j} - \gamma_{i,k}\left(t_{i,j}\right)\right)^2 \\
+ & \sum_{i} \mathbb{1}_{B_{e_i}} \times \log\left(h_i(t_{e_i})\right) + \sum_{i} \log\left(S_i(t_{e_i})\right) \\
- & \log\left(\sigma_{\tilde{g}}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{g}}^2}\left(\tilde{g} - \overline{\tilde{g}}\right)^2 \\
- & \log\left(\sigma_{\tilde{v}_0}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{v}_0}^2}\left(\tilde{v}_0 - \overline{\tilde{v}_0}\right)^2 \\
- & \log\left(\sigma_{\tilde{\nu}}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{\nu}}^2}\left(\tilde{\nu} - \overline{\tilde{\nu}}\right)^2 \\
- & \log\left(\sigma_{\tilde{\rho}}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{\rho}}^2}\left(\tilde{\rho} - \overline{\tilde{\rho}}\right)^2 \\
- & N \log\left(\sigma_\tau\sqrt{2\pi}\right) - \frac{1}{2\sigma_\tau^2}\sum_{i}(\tau_i - \overline{\tau})^2 \\
- & N \log\left(\sigma_\xi\sqrt{2\pi}\right) - \frac{1}{2\sigma_\xi^2}\sum_{i}(\xi_i - \overline{\xi})^2
\end{aligned}
$$

A.1.3 Joint cause-specific Spatiotemporal model

K outcomes and N_s sources and T_e is the time of observation of the event and B_e is the associated boolean whether the event was observed or not (see section 4.2.1), parameters

are defined in section 4.2.3.1.

$$
\begin{aligned}
\log p((y, T_e, B_e), z \mid \theta, \Pi) = & \sum_{i,j,k} -\log\left(\sigma_k\sqrt{2\pi}\right) - \frac{1}{2\sigma_k^2}\left(y_{i,j,k} - \gamma_{i,k}\left(t_{i,j,k}\right)\right)^2 \\
+ & \sum_{i,l} \mathbb{1}_{B_{e_i}=l} \times \log\left(h_{i,l}(t_{e_i})\right) + \sum_{i,l} \log\left(S_{i,l}(t_{e_i})\right) \\
- & \sum_{k} \log\left(\sigma_{\tilde{g}}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{g}}^2}\left(\tilde{g}_k - \overline{\tilde{g}}_k\right)^2 \\
- & \sum_{k} \log\left(\sigma_{\tilde{v}_0}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{v}_0}^2}\left(\tilde{v}_{0,k} - \overline{\tilde{v}}_{0,k}\right)^2 \\
- & \sum_{l} \log\left(\sigma_{\tilde{\nu}}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{\nu}}^2}\left(\tilde{\nu}_l - \overline{\nu_l}\right)^2 \\
- & \sum_{l} \log\left(\sigma_{\tilde{\rho}}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{\rho}}^2}\left(\tilde{\rho}_l - \overline{\rho}_l\right)^2 \\
- & (K-1)N_s \log(\sigma_\beta\sqrt{2\pi}) - \frac{1}{2\sigma_\beta^2}\sum_{o,m}(\beta_{o,m} - \overline{\beta}_{o,m}) \\
- & LN_s \log\left(\sigma_\zeta\sqrt{2\pi}\right) - \frac{1}{2\sigma_\zeta^2}\sum_{l,m}\left(\tilde{\zeta}_{l,m} - \overline{\zeta}_{l,m}\right)^2 \\
- & N \log\left(\sigma_\tau\sqrt{2\pi}\right) - \frac{1}{2\sigma_\tau^2}\sum_{i}(\tau_i - \overline{\tau})^2 \\
- & N \log\left(\sigma_\xi\sqrt{2\pi}\right) - \frac{1}{2\sigma_\xi^2}\sum_{i}(\xi_i - \overline{\xi})^2 \\
- & NN_s \log\left(\sigma_s\sqrt{2\pi}\right) - \frac{1}{2\sigma_s^2}\sum_{i}\sum_{m}^{N_s}(s_{i,m} - \overline{s})^2
\end{aligned}
$$

A.2 Sufficient statistics

The convergence of the Monte-Carlo Markov Chain Stochastic Approximation Expectation-Maximization (MCMC-SAEM) algorithm has been proven in Kuhn and Lavielle, 2004 for models which lie into the curved exponential family. For such a family of distributions, the log-likelihood can be written as:

$$\log\ p(Y, z, \theta, \Pi) = -\Phi(\theta, \Pi) + \langle S(Y, z), f(\theta, \Pi)\rangle + A(Y, z, \Pi)$$

where Φ and f are smooth functions, and S are called the sufficient statistics. The sufficient statistics are to be understood as a summary of the required information from the latent variables z and the observations Y. Our models fall in such a category and sufficient statistics are described below. Note that for the joint models, the same kind of decomposition was done by Lavalley-Morelle et al., 2024.

The idea is to rewrite likelihood in the above form to get sufficient statistics. As a reminder, note that there are N patients indexed by i and each has visits indexed by j.

A.2.1 Longitudinal Spatiotemporal model

Extracted from Koval, 2020 p.175, see notation in section 2.2.1 and parameters in section 2.2.4.1, $t_0 = \overline{\tau}, \overline{\xi} = 0$

$$
\begin{aligned}
\log q(y, z \mid \theta, \Pi) = & -\sum_{i,j,k} \ln(\sigma_k \sqrt{2\pi}) - \langle \underbrace{[\|y_{ij}\|^2]_{ij}}_{S_1(y,z)} \underbrace{-2[y_{ij}^T \gamma_i(t_{i,j})]_{ij}}_{S_2(y,z)} + \underbrace{[\|\gamma_i(t_{i,j})\|^2]_{ij}}_{S_3(y,z)}, \frac{1}{2[\sigma_k^2]_k} \mathbf{1}_{\sum_{i,j,k} 1} \rangle \\
& - KN \ln(\sigma_{\tilde{v}_0} \sqrt{2\pi}) - \sum_{k=1}^{K} \frac{1}{2\sigma_{\tilde{v}_0}^2} \overline{\tilde{v}}_{0k}^2 + \langle \underbrace{[\tilde{v}_{0,k}^2]_k}_{S_4(y,z)}, -\frac{1}{2\sigma_{\tilde{v}_0}^2} \mathbf{1}_K \rangle + \langle \underbrace{[\tilde{v}_{0,k}]_k}_{S_5(y,z)}, \frac{1}{\sigma_{\tilde{v}_0}^2} [\overline{\tilde{v}}_{0k}]_k \rangle \\
& - K \ln(\sigma_{\tilde{g}} \sqrt{2\pi}) - \sum_{k=1}^{K} \frac{1}{2\sigma_{\tilde{g}}^2} \overline{\tilde{g}}_k^2 + \langle \underbrace{[\tilde{g}_k^2]_k}_{S_6(y,z)}, -\frac{1}{2\sigma_{\tilde{g}}^2} \mathbf{1}_K \rangle + \langle \underbrace{[\tilde{g}_k]_k}_{S_7(y,z)}, \frac{1}{\sigma_{\tilde{g}}^2} [\overline{\tilde{g}}_k]_k \rangle \\
& - (K-1) N_s \ln(\sigma_\beta \sqrt{2\pi}) - \sum_{o,m} \frac{1}{2\sigma_\beta^2} \overline{\beta}_{o,m}^2 \\
& + \langle \underbrace{[\beta_{o,m}^2]_o}_{S_8(y,z)}, -\frac{1}{2\sigma_\beta^2} \mathbf{1}_{(K-1)N_s} \rangle + \langle \underbrace{[\beta_{o,m}]_{o,m}}_{S_9(y,z)}, \frac{1}{\sigma_\beta^2} [\overline{\beta}_{o,m}]_{o,m} \rangle \\
& - N \log(\sigma_\tau \sqrt{2\pi}) - \frac{1}{2\sigma_\tau^2} N \overline{\tau}^2 + \langle \underbrace{[\tau_i^2]_i}_{S_{10}(Y,z)}, -\frac{1}{2\sigma_\tau^2} \mathbf{1}_N \rangle + \langle \underbrace{[\tau_i]_i}_{S_{11}(Y,z)}, \frac{1}{\sigma_\tau^2} \overline{\tau} \mathbf{1}_N \rangle \\
& - N \log(\sigma_\xi \sqrt{2\pi}) - \frac{1}{2\sigma_\xi^2} N \overline{\xi}^2 + \langle \underbrace{[\xi_i^2]_i}_{S_{12}(Y,z)}, -\frac{1}{2\sigma_\xi^2} \mathbf{1}_N \rangle + \langle \underbrace{[\xi_i]_i}_{S_{13}(Y,z)}, \frac{1}{\sigma_\xi^2} \overline{\xi} \mathbf{1}_N \rangle \\
& - N N_s \log(\sigma_s \sqrt{2\pi}) - N \sum_{m=1}^{N_s} \frac{1}{2\sigma_s^2} \overline{s_m} + \langle \underbrace{[\tilde{s_{il}}^2]_{il}}_{S_{14}(y,z)}, -\frac{1}{2\sigma_s^2} \mathbf{1}_{N N_s} \rangle + \sum_{m=1}^{N_s} \langle \underbrace{[\tilde{s}_{i,m}]_i}_{S_{15}(y,z)}, \frac{1}{\sigma_s^2} [\overline{s}] \rangle
\end{aligned}
$$

A.2.2 Joint Temporal model

see notation in section 3.2.1 and parameters in section 3.2.3.1,

$$
\begin{aligned}
\log q((y, T_e, B_e), z \mid \theta, \Pi) = & \sum_{i,j} -\log\left(\sigma\sqrt{2\pi}\right) - \langle \underbrace{[\|y_{ij}\|^2]_{ij}}_{S_1(Y,z)} \underbrace{-2[y_{ij}^T \gamma_i(t_{i,j})]_{ij}}_{S_2(Y,z)} + \underbrace{[\|\gamma_i(t_{i,j})\|^2]_{ij}}_{S_3(Y,z)}, \frac{1}{2\sigma^2}\mathbf{1}_{\sum n_i}\rangle \\
& + [\sum_l \mathbb{1}_{B_{e_i}=l} \times \log(h_{i,l}(t_{e_i})) + \sum_{i,l} \log(S_{i,l}(t_{e_i}))]_i, \\
& - \ln(\sigma_{\tilde{g}}\sqrt{2\pi}) - \frac{1}{2\sigma_{\tilde{g}}^2}\overline{\tilde{g}}^2 + \langle \underbrace{[\tilde{g}^2]}_{S_4(Y,z)}, -\frac{1}{2\sigma_{\tilde{g}}^2}\mathbf{1}_1\rangle + \langle \underbrace{[\tilde{g}]}_{S_5(Y,z)}, \frac{1}{\sigma_{\tilde{g}}^2}[\overline{\tilde{g}}]\rangle \\
& - \ln(\sigma_{\tilde{v}_0}\sqrt{2\pi}) - \frac{1}{2\sigma_{\tilde{v}_0}^2}\overline{\tilde{v}_0}^2 + \langle \underbrace{[\tilde{v}_0^2]}_{S_6(Y,z)}, -\frac{1}{2\sigma_{\tilde{v}_0}^2}\mathbf{1}_1\rangle + \langle \underbrace{[\tilde{v}_0]}_{S_7(Y,z)}, \frac{1}{\sigma_{\tilde{v}_0}^2}[\overline{\tilde{v}_0}]\rangle \\
& - \ln(\sigma_{\tilde{\nu}}\sqrt{2\pi}) - \frac{1}{2\sigma_{\tilde{\nu}}^2}\overline{\tilde{\nu}}^2 + \langle \underbrace{[\tilde{\nu}^2]}_{S_8(Y,z)}, -\frac{1}{2\sigma_{\tilde{\nu}}^2}\mathbf{1}_1\rangle + \langle \underbrace{[\tilde{\nu}]}_{S_9(Y,z)}, \frac{1}{\sigma_{\tilde{\nu}}^2}[\overline{\tilde{\nu}}]\rangle \\
& - \ln(\sigma_{\tilde{\rho}}\sqrt{2\pi}) - \frac{1}{2\sigma_{\tilde{\rho}}^2}\overline{\tilde{\rho}}^2 + \langle \underbrace{[\tilde{\rho}^2]}_{S_{10}(Y,z)}, -\frac{1}{2\sigma_{\tilde{\rho}}^2}\mathbf{1}_1\rangle + \langle \underbrace{[\tilde{\rho}]}_{S_{11}(Y,z)}, \frac{1}{\sigma_{\tilde{\rho}}^2}[\overline{\tilde{\rho}}]\rangle \\
& - N\log(\sigma_\tau\sqrt{2\pi}) - \frac{1}{2\sigma_\tau^2}N\overline{\tau}^2 + \langle \underbrace{[\tau_i^2]_i}_{S_{12}(Y,z)}, -\frac{1}{2\sigma_\tau^2}\mathbf{1}_N\rangle + \langle \underbrace{[\tau_i]_i}_{S_{13}(Y,z)}, \frac{1}{\sigma_\tau^2}\overline{\tau}\mathbf{1}_N\rangle \\
& - N\log(\sigma_\xi\sqrt{2\pi}) - \frac{1}{2\sigma_\xi^2}N\overline{\xi}^2 + \langle \underbrace{[\xi_i^2]_i}_{S_{14}(Y,z)}, -\frac{1}{2\sigma_\xi^2}\mathbf{1}_N\rangle + \langle \underbrace{[\xi_i]_i}_{S_{15}(Y,z)}, \frac{1}{\sigma_\xi^2}\overline{\xi}\mathbf{1}_N\rangle
\end{aligned}
$$

A.2.3 Joint cause-specific Spatiotemporal model

see notation in section 4.2.1 and parameters in section 4.2.3.1,

$$
\begin{aligned}
\log q((y, T_e, B_e), z \mid \theta, \Pi) = & -\sum_{i,j,k} \ln(\sigma_k \sqrt{2\pi}) - \langle \underbrace{[\|y_{ij}\|^2]_{ij}}_{S_1(y,z)} \underbrace{-2[y_{ij}^T \gamma_i(t_{i,j})]_{ij}}_{S_2(y,z)} + \underbrace{[\|\gamma_i(t_{i,j})\|^2]_{ij}}_{S_3(y,z)}, \frac{1}{2[\sigma_k^2]_k} \mathbf{1}_{\sum_{i,j,k} 1} \rangle \\
& + [\sum_{i,l} \mathbb{1}_{B_{e_i}=l} \times \log\left(h_{i,l}(t_{e_i})\right) + \sum_{i,l} \log\left(S_{i,l}(t_{e_i})\right)]_i, \\
& - KN \ln(\sigma_{\tilde{v}_0} \sqrt{2\pi}) - \sum_{k=1}^{K} \frac{1}{2\sigma_{\tilde{v}_0}^2} \overline{\tilde{v}_{0k}}^2 + \langle \underbrace{[\tilde{v}_{0,k}^2]_k}_{S_4(y,z)}, -\frac{1}{2\sigma_{\tilde{v}_0}^2} \mathbf{1}_K \rangle + \langle \underbrace{[\tilde{v}_{0,k}]_k}_{S_5(y,z)}, \frac{1}{\sigma_{\tilde{v}_0}^2} [\overline{\tilde{v}_{0k}}]_k \rangle \\
& - K \ln(\sigma_{\tilde{g}} \sqrt{2\pi}) - \sum_{k=1}^{K} \frac{1}{2\sigma_{\tilde{g}}^2} \overline{\tilde{g}}_k^2 + \langle \underbrace{[\tilde{g}_k^2]_k}_{S_6(y,z)}, -\frac{1}{2\sigma_{\tilde{g}}^2} \mathbf{1}_K \rangle + \langle \underbrace{[\tilde{g}_k]_k}_{S_7(y,z)}, \frac{1}{\sigma_{\tilde{g}}^2} [\overline{\tilde{g}}_k]_k \rangle \\
& - (K-1) N_s \ln(\sigma_\beta \sqrt{2\pi}) - \sum_{o,m} \frac{1}{2\sigma_\beta^2} \overline{\beta}_{o,m}^2 \\
& + \langle \underbrace{[\beta_{o,m}^2]_{o,m}}_{S_8(y,z)}, -\frac{1}{2\sigma_\beta^2} \mathbf{1}_{(K-1)N_s} \rangle + \langle \underbrace{[\beta_{o,m}]_{o,m}}_{S_9(y,z)}, \frac{1}{\sigma_\beta^2} [\overline{\beta}_{o,m}]_{o,m} \rangle \\
& - \ln(\sigma_{\tilde{\nu}} \sqrt{2\pi}) - \sum_{l=1}^{L} \frac{1}{2\sigma_{\tilde{\nu}}^2} \overline{\tilde{\nu}_l}^2 + \langle \underbrace{[\tilde{\nu}_l^2]_l}_{S_{10}(Y,z)}, -\frac{1}{2\sigma_{\tilde{\nu}}^2} \mathbf{1}_1 \rangle + \langle \underbrace{[\tilde{\nu}_l]_l}_{S_{11}(Y,z)}, \frac{1}{\sigma_{\tilde{\nu}}^2} [\overline{\tilde{\nu}_l}]_l \rangle \\
& - \ln(\sigma_{\tilde{\rho}} \sqrt{2\pi}) - \sum_{l=1}^{L} \frac{1}{2\sigma_{\tilde{\rho}}^2} \overline{\tilde{\rho}_l}^2 + \langle \underbrace{[\tilde{\rho}_l^2]_l}_{S_{12}(Y,z)}, -\frac{1}{2\sigma_{\tilde{\rho}}^2} \mathbf{1}_1 \rangle + \langle \underbrace{[\tilde{\rho}_l]_l}_{S_{13}(Y,z)}, \frac{1}{\sigma_{\tilde{\rho}}^2} [\overline{\tilde{\rho}_l}]_l \rangle \\
& - LN_s \ln(\sigma_\zeta \sqrt{2\pi}) - \sum_{l,m} \frac{1}{2\sigma_\zeta^2} \overline{\zeta}_{l,m}^2 \\
& + \langle \underbrace{[\zeta_{l,m}^2]_{l,m}}_{S_{14}(y,z)}, -\frac{1}{2\sigma_\zeta^2} \mathbf{1}_{LN_s} \rangle + \langle \underbrace{[\zeta_{l,m}]_{l,m}}_{S_{15}(y,z)}, \frac{1}{\sigma_\zeta^2} [\overline{\zeta}_{l,m}]_{l,m} \rangle \\
& - N \log(\sigma_\tau \sqrt{2\pi}) - \frac{1}{2\sigma_\tau^2} N \overline{\tau}^2 + \langle \underbrace{[\tau_i^2]_i}_{S_{16}(Y,z)}, -\frac{1}{2\sigma_\tau^2} \mathbf{1}_N \rangle + \langle \underbrace{[\tau_i]_i}_{S_{17}(Y,z)}, \frac{1}{\sigma_\tau^2} \overline{\tau} \mathbf{1}_N \rangle \\
& - N \log(\sigma_\xi \sqrt{2\pi}) - \frac{1}{2\sigma_\xi^2} N \overline{\xi}^2 + \langle \underbrace{[\xi_i^2]_i}_{S_{18}(Y,z)}, -\frac{1}{2\sigma_\xi^2} \mathbf{1}_N \rangle + \langle \underbrace{[\xi_i]_i}_{S_{19}(Y,z)}, \frac{1}{\sigma_\xi^2} \overline{\xi} \mathbf{1}_N \rangle \\
& - NN_s \log(\sigma_s \sqrt{2\pi}) - N \sum_{m=1}^{N_s} \frac{1}{2\sigma_s^2} \overline{s_m} \\
& + \langle \underbrace{[\tilde{s_{il}}^2]_{il}}_{S_{20}(y,z)}, -\frac{1}{2\sigma_s^2} \mathbf{1}_{NN_s} \rangle + \sum_{m=1}^{N_s} \langle \underbrace{[\tilde{s}_{im}]_i}_{S_{21}(y,z)}, \frac{1}{\sigma_s^2} [\overline{s}] \rangle
\end{aligned}
$$

A.3 Maximization update rules

To find the update rule of the different parameters, we need to find the new parameter θ that maximizes the log-likelihood. As expressions are convex in θ we can simply derive and look for a critical point. We derive the log-likelihood with respect to each maximised fixed

effect. Note that only maximised fixed effects are updated by a maximization rule, other parameters are latent variables that are sampled. $\overline{\xi}$ is first maximised and then set to 0 and $\overline{s} = 0$ and $\sigma_s = 1$. As a reminder, note that there are N patients indexed by i and that each of them has n_i visits indexed by j. At iteration c, we can use $\tilde{S}^{(c+1)}$ computed with the parameters at iteration c and the formula of $S(Y, z)$ to compute the parameters at iteration $(c+1)$.

A.3.1 Longitudinal Spatiotemporal model

Extracted from Koval, 2020 p.175, see notation in section 2.2.1 and parameters in section 2.2.4.1, and sufficient statistics section A.2

$$
\begin{aligned}
(\sigma^2)^{c+1} &\leftarrow \frac{1}{NK}[\tilde{S}_1^{(c+1)} - 2\tilde{S}_2^{(c+1)} + \tilde{S}_3^{(c+1)}]^T \mathbf{1}_K \\
(\overline{\tilde{v}_{0k}})^{c+1} &\leftarrow \tilde{S}_5^{(c+1)} \\
(\overline{\tilde{g}}_k)^{c+1} &\leftarrow \tilde{S}_7^{(c+1)} \\
(\overline{\beta}_{o,m})^{c+1} &\leftarrow \tilde{S}_9^{(c+1)} \\
(\overline{\tau})^{c+1} &\leftarrow \frac{1}{N}\tilde{S}_{11}^{(c+1)} \\
(\sigma_\tau^2)^{c+1} &\leftarrow \frac{1}{N}[\tilde{S}_{10}^{(c+1)} - 2\overline{\tau}\tilde{S}_{11}^{(c+1)}]^T \mathbf{1}_N + \overline{\tau}^2 \\
(\sigma_\xi^2)^{(c+1)} &\leftarrow \frac{1}{N}[\tilde{S}_{12}^{(c+1)} - 2\overline{\xi}\tilde{S}_{13}^{(c+1)}]^T \mathbf{1}_N + \overline{\xi}^2
\end{aligned}
$$

A.3.2 Joint Temporal model

see notation in section 3.2.1 and parameters in section 3.2.3.1 and sufficient statistics section A.2

$$
\begin{aligned}
(\sigma^2)^{(c+1)} &\leftarrow \frac{1}{N}[\tilde{S}_1^{(c+1)} - 2\tilde{S}_2^{(c+1)} + \tilde{S}_3^{(c)}]^T \mathbf{1}_1 \\
(\overline{\tilde{g}})^{(c+1)} &\leftarrow \tilde{S}_5^{(c+1)} \\
(\overline{\tilde{v}_0})^{(c+1)} &\leftarrow \tilde{S}_7^{(c+1)} \\
(\overline{\tilde{\nu}})^{(c+1)} &\leftarrow \tilde{S}_9^{(c+1)} \\
(\overline{\tilde{\rho}})^{(c+1)} &\leftarrow \tilde{S}_{11}^{(c+1)} \\
(\overline{\tau})^{(c+1)} &\leftarrow \frac{1}{N}\tilde{S}_{13}^{(c+1)} \\
(\sigma_\tau^2)^{(c+1)} &\leftarrow \frac{1}{p}[\tilde{S}_{12}^{(c+1)} - 2\overline{\tau}\tilde{S}_{13}^{(c+1)}]^T \mathbf{1}_N + \overline{\tau}^2 \\
(\overline{\xi})^{(c+1)} &\leftarrow \frac{1}{N}\tilde{S}_{15}^{(c+1)} \\
(\sigma_\xi^2)^{(c+1)} &\leftarrow \frac{1}{N}[\tilde{S}_{14}^{(c+1)} - 2\overline{\xi}\tilde{S}_{15}^{(c+1)}]^T \mathbf{1}_N + \overline{\xi}^2
\end{aligned}
$$

A.3.3 Joint cause-specific Spatiotemporal model

see notation in section 4.2.1 and parameters in section 4.2.3.1, and sufficient statistics section A.2

$$
\begin{aligned}
(\sigma^2)^{(c+1)} &\leftarrow \frac{1}{N}[\tilde{S}_1^{(c+1)} - 2\tilde{S}_2^{(c+1)} + \tilde{S}_3^{(c)}]^T \mathbf{1}_1 \\
(\overline{\tilde{v}_{0k}})^{c+1} &\leftarrow \tilde{S}_5^{(c+1)} \\
(\overline{\tilde{g}}_k)^{c+1} &\leftarrow \tilde{S}_7^{(c+1)} \\
(\overline{\beta}_{o,m})^{c+1} &\leftarrow \tilde{S}_9^{(c+1)} \\
(\overline{\tilde{\nu}}_l)^{(c+1)} &\leftarrow \tilde{S}_{11}^{(c+1)} \\
(\overline{\tilde{\rho}}_l)^{(c+1)} &\leftarrow \tilde{S}_{13}^{(c+1)} \\
(\overline{\zeta}_{l,m})^{c+1} &\leftarrow \tilde{S}_{15}^{(c+1)} \\
(\overline{\tau})^{(c+1)} &\leftarrow \frac{1}{N}\tilde{S}_{17}^{(c+1)} \\
(\sigma_\tau^2)^{(c+1)} &\leftarrow \frac{1}{N}[\tilde{S}_{16}^{(c+1)} - 2\overline{\tau}\tilde{S}_{17}^{(c+1)}]^T \mathbf{1}_N + \overline{\tau}^2 \\
(\overline{\xi})^{(c+1)} &\leftarrow \frac{1}{N}\tilde{S}_{19}^{(c+1)} \\
(\sigma_\xi^2)^{(c+1)} &\leftarrow \frac{1}{N}[\tilde{S}_{18}^{(c+1)} - 2\overline{\xi}\tilde{S}_{19}^{(c+1)}]^T \mathbf{1}_N + \overline{\xi}^2
\end{aligned}
$$

A.4 Jacobian Likelihood

To faster personalisation, gradients are computed for $\xi_i \times \sigma_\xi$ and $\tau_i \times \sigma_\tau$. Thus all the equations must be multiplied by the standard deviation at the end, to get the implemented formulas.

A.4.1 Joint cause-specific Spatiotemporal model

A.4.1.1 Longitudinal data attachment

From likelihood Longitudinal noise is supposed to follow Gaussian law, we have to derive the following quantity per patient i and visit j for the outcome k:

$$
\log p(y_{i,j,k} \mid z, \theta, \Pi) = \quad -\log\left(\sigma_k\sqrt{2\pi}\right) - \tfrac{1}{2\sigma_k^2}\left(y_{i,j,k} - \gamma_{i,k}\left(t_{i,j,k}\right)\right)^2
$$

Jacobian Using the known formula of the derivative of the logistic function, we get:

$$
\begin{aligned}
C_{i,j,k} &= \tfrac{(1+g_k)^2}{g_k}\left[y_{i,j,k} - \gamma_{i,k}\left(t_{i,j,k}\right)\right]\left[\gamma_{i,k}\left(t_{i,j,k}\right)\right]\left[1 - \gamma_{i,k}\left(t_{i,j,k}\right)\right] \\
\frac{\partial \log p(y_{i,j,k} \mid z, \theta, \Pi)}{\partial \xi_i} &= \tfrac{1}{\sigma^2}\left(v_{0,k}\psi_i\left(t_{i,j,k}\right) \times C_{i,j,k}\right) \\
\frac{\partial \log p(y_{i,j,k} \mid z, \theta, \Pi)}{\partial \tau_i} &= -\tfrac{1}{\sigma^2}\left(v_{0,k}e^{\xi_i} \times C_{i,j,k}\right) \\
\frac{\partial \log p(y_{i,j,k} \mid z, \theta, \Pi)}{\partial s_{i,m}} &= -\tfrac{1}{\sigma^2}\left(A_{k,m} \times C_{i,j,k}\right)
\end{aligned}
$$

A.4.1.2 Survival data attachment

From likelihood On the other side, the modelling of the survival process depends on whether the event is observed or not for each patient i:

$$\begin{aligned}\log p(t_{e_i}, B_{e_i} \mid z, \theta, \Pi) = \; & \textstyle\sum_l \mathbb{1}_{B_{e_i}=l} \times \log\left(\frac{\rho_l e^{\xi_i}}{\nu_l}\left(\frac{e^{\xi_i}(t_{e_i}-\tau_i)}{\nu_l}\right)^{\rho_l-1}\exp\left(\textstyle\sum_m \zeta_{l,m}s_{i,m}\right)\right) \\ & - \textstyle\sum_l \left(\frac{e^{\xi_i}(t_{e_i}-\tau_i)}{\nu_l}\right)^{\rho_l}\exp\left(\textstyle\sum_m \zeta_{l,m}s_{i,m}\right)\end{aligned}$$

Jacobian We thus get:

$$\begin{aligned}\frac{\partial \log q(t_{e_i}, B_{e_i} \mid z, \theta, \Pi)}{\partial \xi_i} &= \textstyle\sum_l \mathbb{1}_{B_{e_i}=l}\rho_l - \rho_l \times \left(\frac{\psi_i(t_{e_i})}{\nu_l}\right)^{\rho_l}\exp\left(\textstyle\sum_m \zeta_{l,m}s_{i,m}\right) \\ &= \textstyle\sum_l \mathbb{1}_{B_{e_i}=l}\rho_l + \rho_l \times \log\left(S_{i,l}(t_{e_i})\right) \\ \frac{\partial \log q(t_{e_i}, B_{e_i} \mid z, \theta, \Pi)}{\partial \tau_i} &= \textstyle\sum_l -\frac{(\rho_l-1)}{(t_{e_i}-\tau_i)}\mathbb{1}_{B_{e_i}=l} - \frac{\rho_l e^{\xi_i}}{\nu_l}\left(\frac{\psi_i(t_{e_i})}{\nu_l}\right)^{\rho_l-1}\exp\left(\textstyle\sum_m \zeta_{l,m}s_{i,m}\right) \\ &= \textstyle\sum_l -\frac{(\rho_l-1)}{(t_{e_i}-\tau_i)}\mathbb{1}_{B_{e_i}=l} - h_{i,l}(t_{e_i}) \\ \frac{\partial \log q(t_{e_i}, B_{e_i} \mid z, \theta, \Pi)}{\partial s_{i,m}} &= \textstyle\sum_l \mathbb{1}_{B_{e_i}=l}\zeta_{l,m} - \zeta_{l,m} \times \left(\frac{\psi_i(t_{e_i})}{\nu_l}\right)^{\rho_l}\exp\left(\textstyle\sum_m \zeta_{l,m}s_{i,m}\right) \\ &= \textstyle\sum_l \mathbb{1}_{B_{e_i}=l}\zeta_{l,m} + \zeta_{l,m} \times \log\left(S_{i,l}(t_{e_i})\right)\end{aligned}$$

A.4.1.3 Random effects regularisation

From likelihood

$$\begin{aligned}\log p(z_{re} \mid \theta, \Pi) = & - N\log\left(\sigma_\tau\sqrt{2\pi}\right) - \frac{1}{2\sigma_\tau^2}\sum_i(\tau_i-\overline{\tau})^2 \\ & - N\log\left(\sigma_\xi\sqrt{2\pi}\right) - \frac{1}{2\sigma_\xi^2}\sum_i(\xi_i-\overline{\xi})^2 \\ & - NN_s\log\left(\sigma_s\sqrt{2\pi}\right) - \frac{1}{2\sigma_s^2}\sum_i\sum_m^{N_s}(s_{i,m}-\overline{s})^2\end{aligned}$$

Jacobian We thus get:

$$\begin{aligned}\frac{\partial \log q(\xi_i \mid \theta, \Pi)}{\partial \xi_i} &= \left(\tfrac{\xi_i-\overline{\xi}}{\sigma_\xi}\right) \\ \frac{\partial \log q(\tau_i \mid \theta, \Pi)}{\partial \tau_i} &= \left(\tfrac{\tau_i-\overline{\tau}}{\sigma_\tau}\right) \\ \frac{\partial \log q(s_{i,m} \mid \theta, \Pi)}{\partial s_{i,m}} &= \left(\tfrac{s_{i,m}-\overline{s}}{\sigma_s}\right)\end{aligned}$$

Appendix B

Joint temporal model additional results

In this appendix additional results on the Joint Temporal model are grouped.

Contents

B.1 Simulation study

B.1.1 Data statistics

B.1.2 Posterior effects estimation

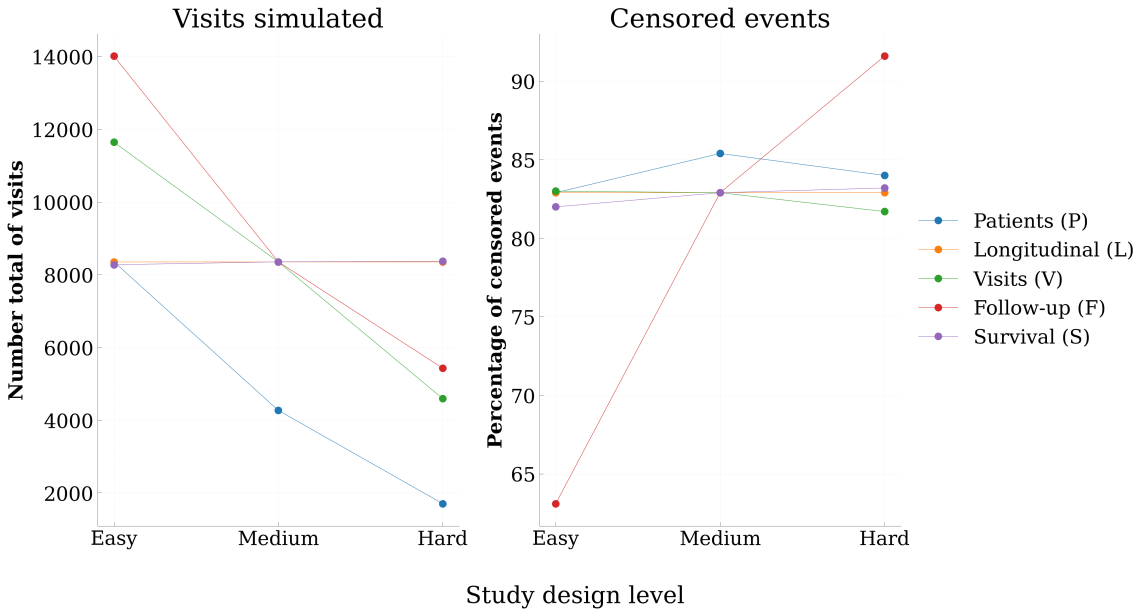


FIGURE B.1: Statistics on simulated data depending on the scenario

Legend: Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. Easy, Medium and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in section 3.3.1.2 and summarised in the table 3.3. Note that the Longitudinal (L) and the Survival (S) scenarios have almost the same number of visits. The Visits (V) and Longitudinal (L) scenarios have almost the same censoring rate for the easiest configuration, and the Survival (S) and Longitudinal (L) scenarios for the most difficult configuration.

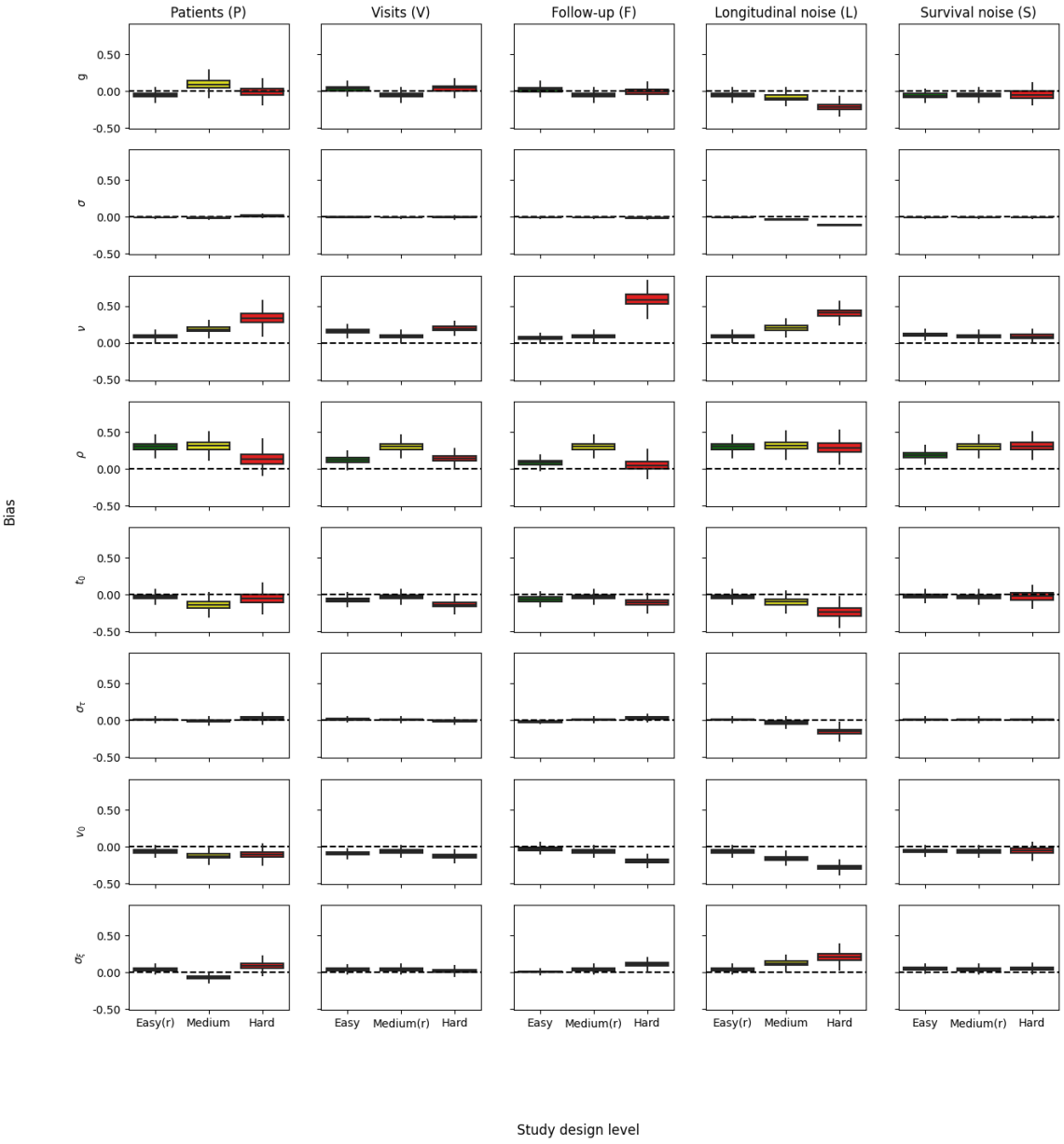


FIGURE B.2: Fixed effects estimation bias from posterior distribution simulated datasets
Legend: In ordinate the normalised bias on the true value of the parameters, in abscissa the different dataset for each scenario.
Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. With (r) that indicates when the study design corresponds to the real-like design. Easy, Medium and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in section 3.3.1.2 and summarised in the table 3.3.
Parameters: g: Curve value at t_0: $\frac{1}{1+g}$, σ: Estimated noise, ν: Scale of the Weibull distribution, ρ: Shape of the Weibull distribution, t_0: Population estimated reference time, σ_τ: estimated reference time standard-deviation, v_0: Speed of the logistic curve, σ_ξ: Individual log-speed factor (the mean is fixed to 0 to ensure identifiability).
The distribution was computed using 30,000 MCMC iterations.

TABLE B.1: Intra-class correlation of random effects from the mean of posterior distribution on simulated datasets

Legend: ξ: the progression log-speed factor, τ: the individual estimated reference time. Intra-class correlations are computed between the random effects estimated and the one simulated. Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. Easy, Medium and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in section 3.3.1.2 and summarised in the table 3.3.

Random Effect	Scenario	Patients (P)		Visits (V)		Follow-up (F)		Longitudinal noise (L)		Survival noise (S)	
		average	CI95%	average	CI95%	average	CI95%	average	CI95%	average	CI95%
ξ	Easy	0.91	[0.90 0.92]	0.93	[0.92 0.93]	0.97	[0.97 0.97]	0.91	[0.90 0.92]	0.91	[0.90 0.92]
	Medium	0.89	[0.87 0.91]	0.91	[0.90 0.92]	0.91	[0.90 0.92]	0.83	[0.81 0.85]	0.91	[0.90 0.92]
	Hard	0.93	[0.91 0.95]	0.90	[0.89 0.91]	0.86	[0.84 0.87]	0.75	[0.72 0.78]	0.91	[0.90 0.92]
τ	Easy	0.96	[0.95 0.96]	0.97	[0.96 0.97]	0.98	[0.97 0.98]	0.96	[0.95 0.96]	0.96	[0.95 0.96]
	Medium	0.94	[0.92 0.95]	0.96	[0.95 0.96]	0.96	[0.95 0.96]	0.88	[0.86 0.89]	0.96	[0.95 0.96]
	Hard	0.96	[0.95 0.97]	0.94	[0.93 0.94]	0.92	[0.91 0.93]	0.66	[0.62 0.70]	0.96	[0.95 0.96]

B.1.3 Attenuation bias correction

TABLE B.2: Clinically meaningful estimated parameters signed bias on simulated data for the Joint Temporal model

Legend: CMEP: Clinically meaningful estimated parameters, bias: mean (SD) computed from the posterior distribution, corrected bias: the mean extracted from the Robbin Monroe algorithm corrected for attenuation bias, truth: the value used for data simulation. Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. Easy, Medium and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in section 3.3.1.2 and summarised in the table 3.3.

CMEP	Scenario	Patients (P)			Visits (V)			Follow-up (F)			Longitudinal noise (L)			Survival noise (S)		
		bias	corrected bias	truth	bias	corrected bias	truth	bias	corrected bias	truth	bias	corrected bias	truth	bias	corrected bias	truth
median	Easy	0.40 (0.07)	0.00	4.50	0.50 (0.08)	0.13	4.47	0.18 (0.05)	0.05	4.40	0.40 (0.07)	0.00	4.50	0.38 (0.06)	0.02	4.18
	Medium	0.59 (0.13)	0.09	4.76	0.40 (0.07)	0.00	4.50	0.40 (0.07)	0.00	4.50	0.71 (0.11)	0.07	4.50	0.40 (0.07)	0.00	4.50
	Hard	1.11 (0.21)	0.58	4.34	0.56 (0.09)	0.14	4.42	1.88 (0.24)	1.08	4.49	1.22 (0.15)	0.43	4.50	0.45 (0.08)	0.02	4.61
growth	Easy	1.37 (0.14)	0.45	-13.19	1.04 (0.13)	0.69	-13.22	0.25 (0.08)	0.50	-13.25	1.37 (0.14)	0.45	-13.19	1.37 (0.14)	0.50	-13.09
	Medium	0.89 (0.19)	0.76	-12.47	1.37 (0.14)	0.45	-13.19	1.37 (0.14)	0.45	-13.19	2.92 (0.21)	-0.59	-13.19	1.37 (0.14)	0.45	-13.19
	Hard	1.68 (0.32)	0.37	-13.49	1.46 (0.15)	0.43	-13.39	2.75 (0.19)	-0.26	-13.36	5.26 (0.24)	-3.51	-13.19	1.32 (0.14)	0.56	-13.22
midpoint	Easy	0.10 (0.02)	-0.06	2.93	0.08 (0.02)	-0.07	2.90	-0.02 (0.01)	-0.05	2.83	0.10 (0.02)	-0.06	2.93	0.11 (0.02)	-0.05	2.94
	Medium	0.04 (0.03)	-0.13	3.09	0.10 (0.02)	-0.06	2.93	0.10 (0.02)	-0.06	2.93	0.25 (0.03)	-0.02	2.93	0.10 (0.02)	-0.06	2.93
	Hard	0.17 (0.04)	0.01	2.80	0.08 (0.02)	-0.07	2.87	0.29 (0.03)	-0.01	2.93	0.46 (0.05)	0.09	2.93	0.10 (0.02)	-0.06	2.91

TABLE B.3: Intra-class correlation of random effects from personalization on simulated datasets for the Joint Temporal model

Legend: ξ: the progression log-speed factor, τ: the individual time-shit, . Intra-class correlations are computed between the random effects estimated and the one simulated. Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. Easy, Medium and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in section 3.3.1.2 and summarised in the table 3.3.

Random Effect	Scenario	Patients (P)		Visits (V)		Follow-up (F)		Longitudinal noise (L)		Survival noise (S)	
		average	CI95%	average	CI95%	average	CI95%	average	CI95%	average	CI95%
ξ	Easy	0.87	[0.86 0.89]	0.89	[0.87 0.90]	0.96	[0.95 0.96]	0.87	[0.86 0.89]	0.87	[0.85 0.88]
	Medium	0.83	[0.80 0.86]	0.87	[0.86 0.89]	0.87	[0.86 0.89]	0.75	[0.71 0.78]	0.87	[0.86 0.89]
	Hard	0.90	[0.87 0.93]	0.86	[0.85 0.88]	0.79	[0.76 0.81]	0.67	[0.63 0.71]	0.87	[0.86 0.89]
τ	Easy	0.97	[0.96 0.97]	0.97	[0.97 0.98]	0.98	[0.98 0.98]	0.97	[0.96 0.97]	0.97	[0.96 0.97]
	Medium	0.95	[0.94 0.96]	0.97	[0.96 0.97]	0.97	[0.96 0.97]	0.91	[0.90 0.92]	0.97	[0.96 0.97]
	Hard	0.96	[0.95 0.97]	0.96	[0.95 0.96]	0.95	[0.94 0.96]	0.71	[0.67 0.75]	0.97	[0.96 0.97]

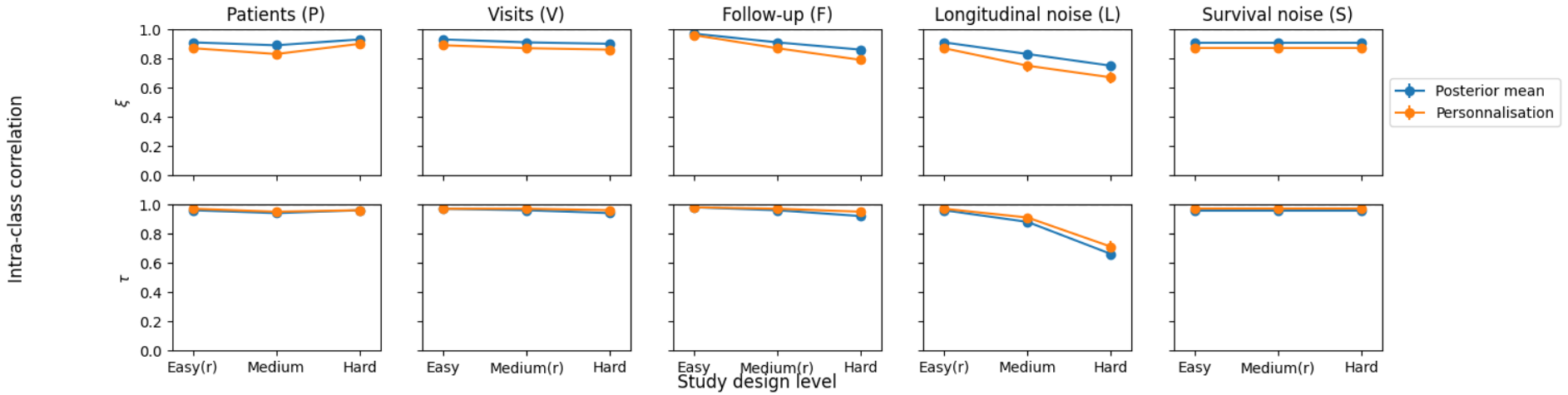


FIGURE B.3: Intra-class correlation of random effects on simulated datasets

Legend: Results are presented with mean with CI95%.
In ordinate, the two random effects: τ the individual time-shit and ξ the progression log-speed factor. Intra-class correlations are computed between the random effects estimated and the one simulated.
In abscissa the different dataset for each scenario: Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. Easy, Medium and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in section 3.3.1.2 and summarised in the table 3.3. Datasets with (r) correspond to the real-like dataset.

B.1.4 Benchmark

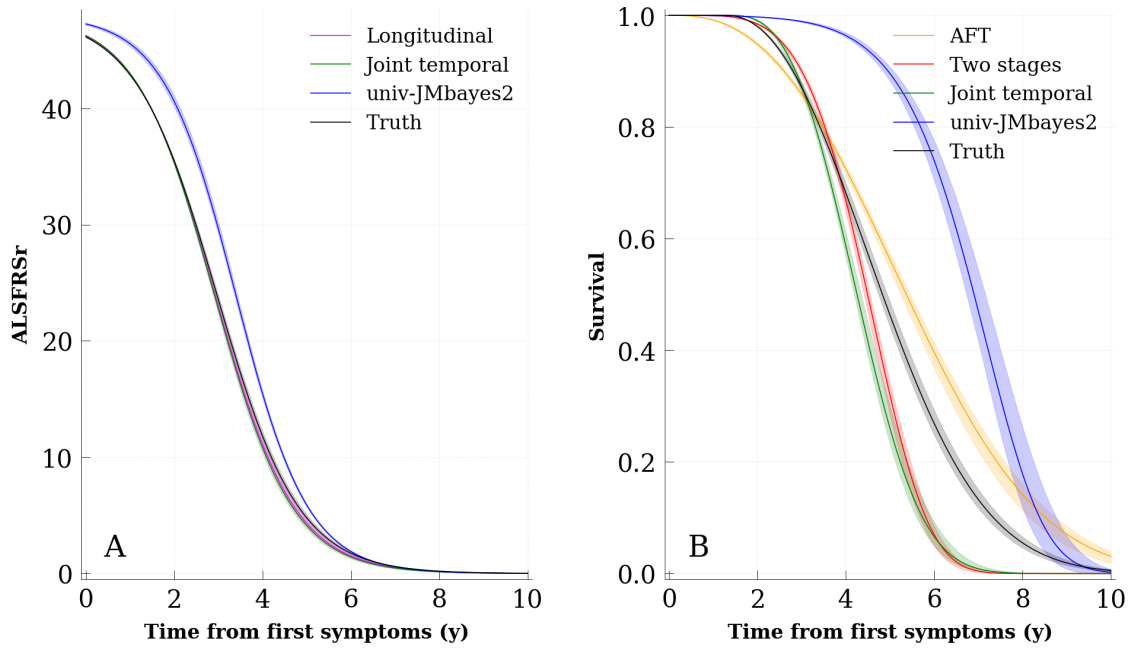


FIGURE B.4: Average patient curve of the Joint Temporal model and reference models on real-like dataset

Legend: Results are presented with mean over the 10-fold cross-validation and the maximum and minimum variation for each model.
Panel A: *Longitudinal: Spatiotemporal longitudinal model, Joint Temporal: the Joint Temporal model, univ-JMbayes2: joint model with shared random effects, Truth: real average patient that enabled data simulation,*
Panel B: *AFT: Accelerated Failure Time model, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, Joint Temporal: the Joint Temporal model, univ-JMbayes2: joint model with shared random effects, Truth: real average patient that enabled data simulation.*
Note that for the longitudinal process, the curves of the Truth, the Longitudinal and Joint Temporal models are superimposed.

B.2 Real data study

B.2.1 Benchmark

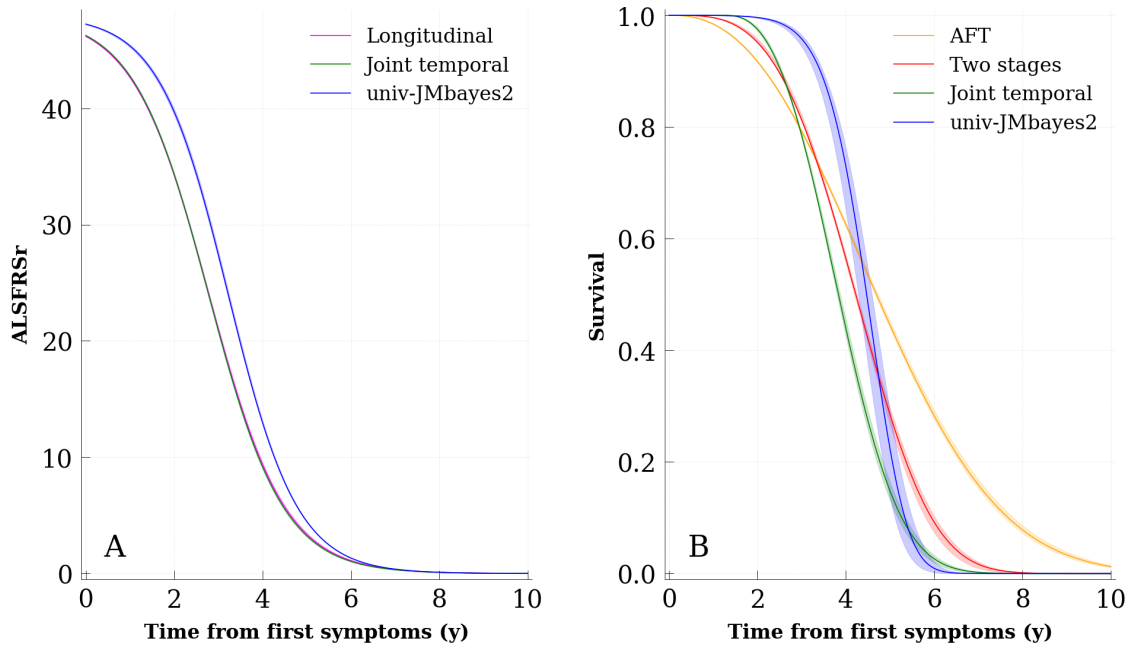


FIGURE B.5: Average patient curve of the Joint Temporal model and reference models on PRO-ACT dataset

Legend: Results are presented with mean over the 10-fold cross-validation and the maximum and minimum variation for each model.

Panel A: *Longitudinal: Spatiotemporal longitudinal model, Joint Temporal: the Joint Temporal model, univ-JMbayes2: joint model with shared random effects,*

Panel B: *AFT: Accelerated Failure Time model, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, Joint Temporal: the Joint Temporal model, univ-JMbayes2: joint model with shared random effects.*

Note that for the longitudinal process, the curves of the Longitudinal and Joint Temporal models are superimposed.

Appendix C

Joint cause-specific Spatiotemporal model

Computation method for WAIC computation and some additional results for the Joint cause-specific model are grouped.

Contents

C.1 WAIC

The WAIC was first defined by Watanabe, 2010 to compare the different Bayesian models. We used the version of Vehtari, Gelman, and Gabry, 2017 with $\beta = 1$ and multiplied by - 2 to be on the deviance scale. To compute it, the probability observation knowing the parameters should be computed for each iteration.

As we used mixed-effects models, two types of likelihood could be computed for each patient i, with $\hat{\theta}$ the model parameters estimated at the end of the computation and Π the hyperparameters :

- the conditional likelihood for iteration c:

$$p((y_i, t_{e_i}, B_{e_i}) \mid \hat{z}_{re_i}, z^c_{fe}, \hat{\theta}, \Pi)$$

  with $\hat{z}_{re_i}$ the mean random effects estimated over the iterations,

- the marginal likelihood for iteration c:

$$p((y_i, t_{e_i}, B_{e_i}) \mid z^c_{fe}, \hat{\theta}, \Pi) = \int p((y_i, t_{e_i}, B_{e_i}), z_{re_i} \mid z^c_{fe}, \hat{\theta}, \Pi) dz_{re_i}$$

The marginal likelihood is more robust (Millar, 2018) but is harder to compute as the integral must be estimated. It is usually estimated using Laplace's approximation which corresponds to a Taylor expansion (Daxberger et al., 2021):

$$\log(p((y_i, t_{e_i}, B_{e_i}) \mid z_{fe}^c, \hat{\theta}, \Pi)) = \log(p((y_i, t_{e_i}, B_{e_i}), \hat{z}_{re_i} \mid z_{fe}^c, \hat{\theta}, \Pi)) + \frac{1}{2}\log \mid S \mid + \frac{D}{2}\log(2\pi)$$

with D the number of random effects and S Hessian (the second derivative) of random effects at point $\hat{z}_{re_i}$. For computational reasons, in the thesis, we estimated only the conditional version of the WAIC.

C.2 Simulation study

C.2.1 Precision of the parameters estimated

C.2.1.1 Real-like dataset additional results

TABLE C.1: Intraclass correlation of random effects of the Joint cause-specific Spatiotemporal model estimated on the Real-like dataset

Legend: Mean intraclass correlation with 95% Confidence Interval estimated using the mean of each individual posterior

Random effects			Intraclass correlation
Log-speed factor		ξ_i	0.86 [0.85 0.88]
Individual estimated reference time		τ_i	0.94 [0.93 0.94]
Space shifts	Bulbar	$w_{i,0}$	0.98 [0.97 0.98]
	Fine motor	$w_{i,1}$	0.96 [0.95 0.96]
	Gross motor	$w_{i,2}$	0.97 [0.97 0.97]
	Total	$w_{i,3}$	0.96 [0.96 0.97]
Survival shifts	VNI	$u_{i,0}$	0.94 [0.93 0.95]
	Death	$u_{i,1}$	0.83 [0.81 0.85]

C.2.1.2 No-link dataset additional results

TABLE C.2: Bias of the estimated parameters of the Joint cause-specific Spatiotemporal model on the No-link simulated dataset

Legend: For each parameter of the model, the value used for simulation is reported with the value of the mean of the stabilized posterior and the associated Credibility Interval at 95% computed on the last 10,000 iterations of the stabilized MCMC-SAEM: (*) the simulated value is in the 95% credibility interval, (<) the simulated value is underestimated, (>) the simulated value is overestimated. Bias: difference to the true parameters divided by the true parameters,-: put when the bias could not be computed as the true parameter equal to 0, ×: no true value for this parameter exist (the number of sources was overestimated due to the invalid shared latent age hypothesis), $\overline{\xi}, \overline{s}, \sigma_s$ *parameters are not present as they are fixed by the model (*$\overline{\xi} = 0, \overline{s} = 0, \sigma_s = 1$*) and* $t_0 = \overline{\tau}$

	Parameters name		Simulated		Estimated	Bias (%)
Distribution of random effects	Estimated reference time (mean)	t_0	1.115	*	1.152 [1.09, 1.215]	3.3 [-2.3, 9.0]
	Estimated reference time (std)	σ_τ	0.988	*	0.99 [0.967, 1.013]	0.2 [-2.1, 2.5]
	Individual log-speed factor (std)	σ_ξ	0.768	*	0.793 [0.767, 0.821]	3.3 [-0.1, 6.9]
Longitudinal fixed effects	Curve values at t_0: $\frac{1}{1+g}$ (g_k)	g_0	13.958	<	12.553 [11.947, 13.434]	-10.1 [-14.4, -3.8]
		g_1	5.316	<	4.438 [3.949, 4.74]	-16.5 [-25.7, -10.8]
		g_2	3.993	<	3.42 [3.225, 3.631]	-14.3 [-19.2, -9.1]
		g_3	5.704	<	5.296 [5.119, 5.475]	-7.2 [-10.3, -4.0]
	Speed of the logistic curves ($v_{0,k}$)	$v_{0,0}$	0.069	*	0.071 [0.067, 0.075]	2.9 [-2.7, 7.8]
		$v_{0,1}$	0.188	*	0.198 [0.187, 0.213]	4.9 [-0.5, 13.0]
		$v_{0,2}$	0.198	*	0.202 [0.193, 0.21]	2.0 [-2.7, 6.1]
		$v_{0,3}$	0.113	*	0.113 [0.109, 0.117]	0.4 [-3.0, 4.0]
	Mixing matrix ($A_{k,m}$)	$A_{0,0}$	0.06	<	0.025 [0.016, 0.036]	-58.3 [-73.3, -40.8]
		$A_{0,1}$	0.059	<	-0.038 [-0.045, -0.033]	-164.7 [-176.0, -155.9]
		$A_{0,2}$	×	×	-0.071 [-0.077, -0.065]	×
		$A_{1,0}$	-0.1	>	-0.053 [-0.06, -0.043]	47.3 [39.8, 56.6]
		$A_{1,1}$	0.006	>	0.095 [0.088, 0.108]	1482.9 [1362.7, 1692.1]
		$A_{1,2}$	×	×	0.016 [0.009, 0.025]	×
		$A_{2,0}$	0.0	*	0.008 [-0.009, 0.028]	-
		$A_{2,1}$	-0.14	>	-0.03 [-0.04, -0.014]	78.7 [71.2, 89.8]
		$A_{2,2}$	×	×	0.144 [0.138, 0.151]	×
		$A_{3,0}$	0.01	<	0.004 [0.002, 0.007]	-55.0 [-77.2, -30.8]
		$A_{3,1}$	0.0	<	-0.008 [-0.01, -0.007]	-
		$A_{3,2}$	×	×	-0.002 [-0.003, -0.0]	×
	Estimated noises (σ_k)	σ_0	0.065	<	0.059 [0.058, 0.059]	-9.9 [-10.5, -9.3]
		σ_1	0.076	<	0.072 [0.071, 0.072]	-5.9 [-6.6, -5.1]
		σ_2	0.074	<	0.069 [0.069, 0.07]	-6.5 [-7.1, -5.7]
		σ_3	0.036	*	0.036 [0.036, 0.037]	0.0 [-0.8, 0.8]
Survival fixed effects	Weibull scale (ν_l)	ν_0	2.792	>	13.368 [11.838, 15.678]	378.8 [324.0, 461.5]
		ν_1	3.59	>	91.026 [75.775, 115.442]	2435.7 [2010.9, 3115.9]
	Weibull shape (ρ_l)	ρ_0	1.7	<	0.848 [0.789, 0.908]	-50.1 [-53.6, -46.6]
		ρ_1	2.8	<	0.864 [0.783, 0.989]	-69.1 [-72.0, -64.7]
	HR Coefficients ($\zeta_{l,m}$)	$\zeta_{0,0}$	-0.09	*	-0.115 [-0.27, 0.001]	-27.4 [-199.7, 100.9]
		$\zeta_{0,1}$	-0.1	*	0.021 [-0.155, 0.179]	121.0 [-55.3, 279.3]
		$\zeta_{0,2}$	×	×	-0.117 [-0.223, -0.014]	×
		$\zeta_{1,0}$	0.09	>	0.979 [0.665, 1.355]	987.9 [638.6, 1405.5]
		$\zeta_{1,1}$	0.0	<	-0.541 [-0.73, -0.365]	-
		$\zeta_{1,2}$	×	×	-0.078 [-0.274, 0.089]	×

TABLE C.3: Intraclass correlation of random effects of the Joint cause-specific Spatiotemporal model estimated on the No-link dataset

Legend: Mean intraclass correlation with 95% Confidence Interval estimated using the mean of each individual posterior

Random effects			Intraclass correlation
Log-speed factor		ξ_i	0.90 [0.89 0.91]
Individual estimated reference time		τ_i	0.94 [0.93 0.94]
Space shifts	Bulbar	$w_{i,1}$	0.97 [0.96 0.97]
	Fine motor	$w_{i,2}$	0.94 [0.93 0.95]
	Gross motor	$w_{i,3}$	0.96 [0.95 0.96]
	Total	$w_{i,4}$	0.94 [0.93 0.95]
Survival shifts	VNI	$u_{i,1}$	-0.91 [-0.92 -0.9]
	Death	$u_{i,2}$	0.01 [-0.05 0.07]

C.2.2 Shared latent age hypothesis

C.2.2.1 Real-like dataset additional results

TABLE C.4: Comparison of the parameters estimated by the Joint cause-specific Spatiotemporal model and the Longitudinal Spatiotemporal model on the Real-like dataset

Legend: For each parameter of the model, the value used for simulation is reported with the value of the mean of the stabilized posterior and the associated Credibility Interval at 95% computed on the last 10,000 iterations of the stabilized MCMC-SAEM. Comparison: test if the credibility interval are disjoint (*) the 95% credibility interval overlap, (<) the credibility interval of the Joint model is below the one of the longitudinal model, (>) the credibility interval of the Joint model is larger than the one of the longitudinal model, $\overline{\xi}, \overline{s}, \sigma_s$ *parameters are not present as they are fixed by the model (*$\overline{\xi} = 0, \overline{s} = 0, \sigma_s = 1$*) and* $t_0 = \overline{\tau}$

	Parameters name		Simulated	Joint Spatiotemporal	Longitudinal Spatiotemporal	Comparison
Distribution of random effects	Estimated reference time (mean)	t_0	1.36	1.381 [1.336, 1.421]	1.419 [1.415, 1.429]	*
	Estimated reference time (std)	σ_τ	1.062	1.058 [1.038, 1.080]	1.078 [1.073, 1.088]	*
	Individual log-speed factor (std)	σ_ξ	0.725	0.749 [0.722, 0.780]	0.759 [0.755, 0.767]	*
Longitudinal fixed effects	Curve values at t_0: $\frac{1}{1+g}$ (g_k)	g_0	13.958	13.002 [12.344, 13.712]	12.443 [12.22, 12.675]	*
		g_1	5.316	4.549 [4.229, 4.850]	4.26 [4.185, 4.338]	*
		g_2	3.993	3.639 [3.485, 3.828]	3.398 [3.316, 3.466]	>
		g_3	5.704	5.428 [5.308, 5.534]	5.198 [5.133, 5.258]	>
	Speed of the logistic curves ($v_{0,k}$)	$v_{0,0}$	0.062	0.059 [0.056, 0.061]	0.057 [0.056, 0.059]	*
		$v_{0,1}$	0.167	0.165 [0.157, 0.173]	0.162 [0.159, 0.166]	*
		$v_{0,2}$	0.176	0.167 [0.161, 0.173]	0.164 [0.16, 0.167]	*
		$v_{0,3}$	0.1	0.094 [0.091, 0.096]	0.091 [0.089, 0.093]	*
	Mixing matrix ($A_{k,m}$)	$A_{0,0}$	0.06	0.059 [0.053, 0.067]	0.062 [0.052, 0.071]	*
		$A_{0,1}$	0.059	0.059 [0.051, 0.066]	0.061 [0.051, 0.073]	*
		$A_{1,0}$	-0.1	-0.107 [-0.115, -0.101]	-0.112 [-0.118, -0.106]	*
		$A_{1,1}$	0.006	0.009 [-0.003, 0.023]	0.01 [-0.011, 0.029]	*
		$A_{2,0}$	0.0	-0.005 [-0.024, 0.007]	-0.007 [-0.029, 0.015]	*
		$A_{2,1}$	-0.14	-0.147 [-0.154, -0.138]	-0.151 [-0.158, -0.144]	*
		$A_{3,0}$	0.01	0.01 [0.009, 0.011]	0.011 [0.009, 0.012]	*
		$A_{3,1}$	-0.0	-0.0 [-0.002, 0.001]	-0.0 [-0.003, 0.002]	*
	Estimated noises (σ_k)	σ_0	0.065	0.058 [0.057, 0.058]	0.058 [0.058, 0.058]	*
		σ_1	0.076	0.073 [0.072, 0.074]	0.073 [0.073, 0.073]	*
		σ_2	0.074	0.072 [0.071, 0.073]	0.072 [0.072, 0.072]	*
		σ_3	0.036	0.036 [0.036, 0.037]	0.036 [0.036, 0.036]	*

C.2.2.2 No-link dataset additional results

TABLE C.5: Comparison of the parameters estimated by the Joint cause-specific Spatiotemporal model and the Longitudinal Spatiotemporal model on the No-link dataset

Legend: For each parameter of the model, the value used for simulation is reported with the value of the mean of the stabilized posterior and the associated Credibility Interval at 95% computed on the last 10,000 iterations of the stabilized MCMC-SAEM, ×: no true value for this parameter exists (the number of sources was overestimated due to the invalid shared latent age hypothesis),Comparison: test if the credibility interval are disjoint (*) the 95% credibility interval overlap, (<) the credibility interval of the Joint model is below the one of the longitudinal model, (>) the credibility interval of the Joint model is larger than the one of the longitudinal model, $\overline{\xi}, \overline{s}, \sigma_s$ *parameters are not present as they are fixed by the model (*$\overline{\xi} = 0, \overline{s} = 0, \sigma_s = 1$*) and* $t_0 = \overline{\tau}$

	Parameters name		Simulated	Joint Spatiotemporal	Longitudinal Spatiotemporal	Comparison
Distribution of random effects	Estimated reference time (mean)	t_0	1.115	1.152 [1.09, 1.215]	1.098 [1.087, 1.107]	*
	Estimated reference time (std)	σ_τ	0.988	0.99 [0.967, 1.013]	0.999 [0.991, 1.003]	*
	Individual log-speed factor (std)	σ_ξ	0.768	0.793 [0.767, 0.821]	0.814 [0.81, 0.818]	*
Longitudinal fixed effects	Curve values at t_0: $\frac{1}{1+g}$ (g_k)	g_0	13.958	12.553 [11.947, 13.434]	12.399 [12.137, 12.671]	*
		g_1	5.316	4.438 [3.949, 4.74]	4.679 [4.599, 4.76]	*
		g_2	3.993	3.42 [3.225, 3.631]	3.628 [3.554, 3.705]	*
		g_3	5.704	5.296 [5.119, 5.475]	5.421 [5.355, 5.486]	*
	Speed of the logistic curves ($v_{0,k}$)	$v_{0,0}$	0.069	0.071 [0.067, 0.075]	0.068 [0.066, 0.07]	*
		$v_{0,1}$	0.188	0.198 [0.187, 0.213]	0.181 [0.177, 0.186]	*
		$v_{0,2}$	0.198	0.202 [0.193, 0.21]	0.186 [0.182, 0.191]	>
		$v_{0,3}$	0.113	0.113 [0.109, 0.117]	0.106 [0.103, 0.108]	>
	Mixing matrix ($A_{k,m}$)	$A_{0,0}$	0.06	0.025 [0.016, 0.036]	0.045 [0.039, 0.05]	<
		$A_{0,1}$	0.059	-0.038 [-0.045, -0.033]	-0.047 [-0.054, -0.042]	*
		$A_{0,2}$	×	-0.071 [-0.077, -0.065]	-0.056 [-0.061, -0.05]	<
		$A_{1,0}$	-0.1	-0.053 [-0.06, -0.043]	-0.077 [-0.085, -0.071]	>
		$A_{1,1}$	0.006	0.095 [0.088, 0.108]	0.071 [0.064, 0.081]	>
		$A_{1,2}$	×	0.016 [0.009, 0.025]	-0.016 [-0.025, -0.008]	>
		$A_{2,0}$	0.0	0.008 [-0.009, 0.028]	-0.003 [-0.015, 0.01]	*
		$A_{2,1}$	-0.14	-0.03 [-0.04, -0.014]	0.016 [0.008, 0.026]	<
		$A_{2,2}$	×	0.144 [0.138, 0.151]	0.144 [0.137, 0.151]	*
		$A_{3,0}$	0.01	0.004 [0.002, 0.007]	0.006 [0.004, 0.009]	*
		$A_{3,1}$	-0.0	-0.008 [-0.01, -0.007]	-0.007 [-0.009, -0.005]	*
		$A_{3,2}$	×	-0.002 [-0.003, -0.0]	0.001 [-0.001, 0.002]	*
	Estimated noises (σ_k)	σ_0	0.065	0.059 [0.058, 0.059]	0.059 [0.059, 0.059]	*
		σ_1	0.076	0.072 [0.071, 0.072]	0.072 [0.072, 0.072]	*
		σ_2	0.074	0.069 [0.069, 0.07]	0.069 [0.069, 0.069]	*
		σ_3	0.036	0.036 [0.036, 0.037]	0.036 [0.036, 0.036]	*

C.3 Application study

TABLE C.6: Description of the parameters estimated by the Joint cause-specific Spatiotemporal model for the Analysis dataset

Legend: *m indexed the dimension of the ordering of longitudinal outcomes (spatial variability), here with dimension two (two sources). Individual log-speed factor mean $\overline{\xi}$, individual spatial variability mean and standard deviation $\overline{s}, \sigma_s$ parameters are not present as they are fixed by the model ($\overline{\xi} = 0, \overline{s} = 0, \sigma_s = 1$) and $t_0 = \overline{\tau}$*

Distribution of random effects			
Estimated reference time (mean)		t_0	
Estimated reference time (std)		σ_τ	
Individual log-speed factor (std)		σ_ξ	
Longitudinal fixed effects			
	Bulbar	Fine motor	Gros motor
Curve values at t_0: $\frac{1}{1+g}$ (g_k)	g_0	g_1	g_2
Speed of the logistic curves ($v_{0,k}$)	$v_{0,0}$	$v_{0,1}$	$v_{0,2}$
Estimated noises (σ_k)	σ_0	σ_1	σ_2
Mixing matrix ($A_{k,m}$)	$A_{0,0}$	$A_{1,0}$	$A_{2,0}$
	$A_{0,1}$	$A_{1,1}$	$A_{2,1}$
Survival fixed effects			
	NIV initiation		Death
Weibull scale (ν_l)	ν_0		ν_1
Weibull shape (ρ_l)	ρ_0		ρ_1
HR Coefficients ($\zeta_{l,m}$)	$\zeta_{0,0}$		$\zeta_{1,0}$
	$\zeta_{0,1}$		$\zeta_{1,1}$

TABLE C.7: Estimated parameters of the Joint cause-specific Spatiotemporal model on the Analysis dataset

Legend: For each parameter of the model, the value used for simulation is reported with the value of the mean of the stabilized posterior and the associated Credibility Interval at 95% computed on the last 10,000 iterations of the stabilized MCMC-SAEM: (*) the simulated value is in the 95% credibility interval, (<) the simulated value is underestimated, (>) the simulated value is overestimated. Bias: difference to the true parameters divided by the true parameters, $\overline{\xi}, \overline{s}, \sigma_s$ *parameters are not present as they are fixed by the model (*$\overline{\xi} = 0, \overline{s} = 0, \sigma_s = 1$*) and* $t_0 = \overline{\tau}$

	Parameters name		Estimated
Distribution of random effects	Estimated reference time (mean)	t_0	0.889 [0.883, 0.902]
	Estimated reference time (std)	σ_τ	0.974 [0.959, 0.986]
	Individual log-speed factor (std)	σ_ξ	0.782 [0.772, 0.796]
Longitudinal fixed effects	Curve values at t_0: $\frac{1}{1+g}$ (g_k)	g_0	23.749 [23.053, 24.195]
		g_1	9.571 [9.393, 9.735]
		g_2	6.577 [6.376, 6.706]
	Speed of the logistic curves $(v_{0,k})$	$v_{0,0}$	0.042 [0.041, 0.043]
		$v_{0,1}$	0.125 [0.122, 0.127]
		$v_{0,2}$	0.144 [0.142, 0.148]
	Mixing matrix $(A_{k,m})$	$A_{0,0}$	-0.023 [-0.025, -0.021]
		$A_{0,1}$	-0.059 [-0.061, -0.057]
		$A_{1,0}$	-0.026 [-0.029, -0.023]
		$A_{1,1}$	0.066 [0.064, 0.069]
		$A_{2,0}$	0.098 [0.096, 0.102]
		$A_{2,1}$	0.047 [0.044, 0.051]
	Estimated noises (σ_k)	σ_0	0.063 [0.063, 0.063]
		σ_1	0.075 [0.075, 0.075]
		σ_2	0.074 [0.074, 0.074]
Survival fixed effects	Weibull scale (ν_l)	ν_0	3.702 [3.622, 3.799]
		ν_1	4.223 [4.119, 4.31]
	Weibull shape (ρ_l)	ρ_0	1.862 [1.81, 1.906]
		ρ_1	3.338 [3.256, 3.472]
	HR Coefficients $(\zeta_{l,m})$	$\zeta_{0,0}$	0.17 [0.147, 0.215]
		$\zeta_{0,1}$	-0.049 [-0.078, -0.02]
		$\zeta_{1,0}$	0.275 [0.25, 0.338]
		$\zeta_{1,1}$	-0.154 [-0.178, -0.128]

Appendix D

Software additional figures

In this appendix additional results are grouped.

Contents

D.1 Extraction of events from standard of care medical terms

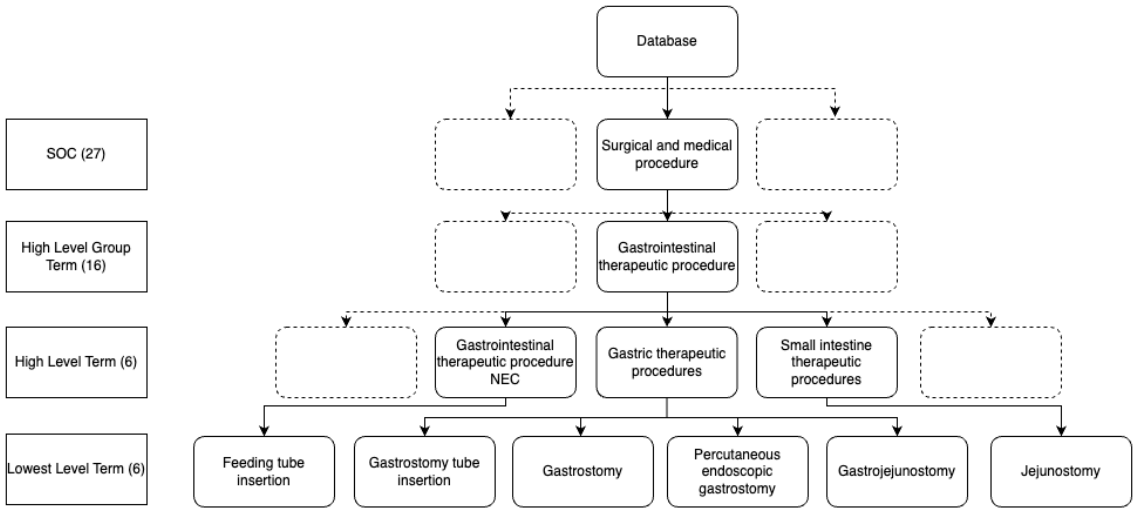


FIGURE D.1: Medical intervention terms related to gastrostomy using Medical Dictionary for Regulatory Activities (MDRA)

Legend: SOC: Standard Of Care, each square represents a level in the medical intervention terms tree, the numbers in parentheses are the number of possible terms at each level of the tree, not selected levels and terms are represented with dashed

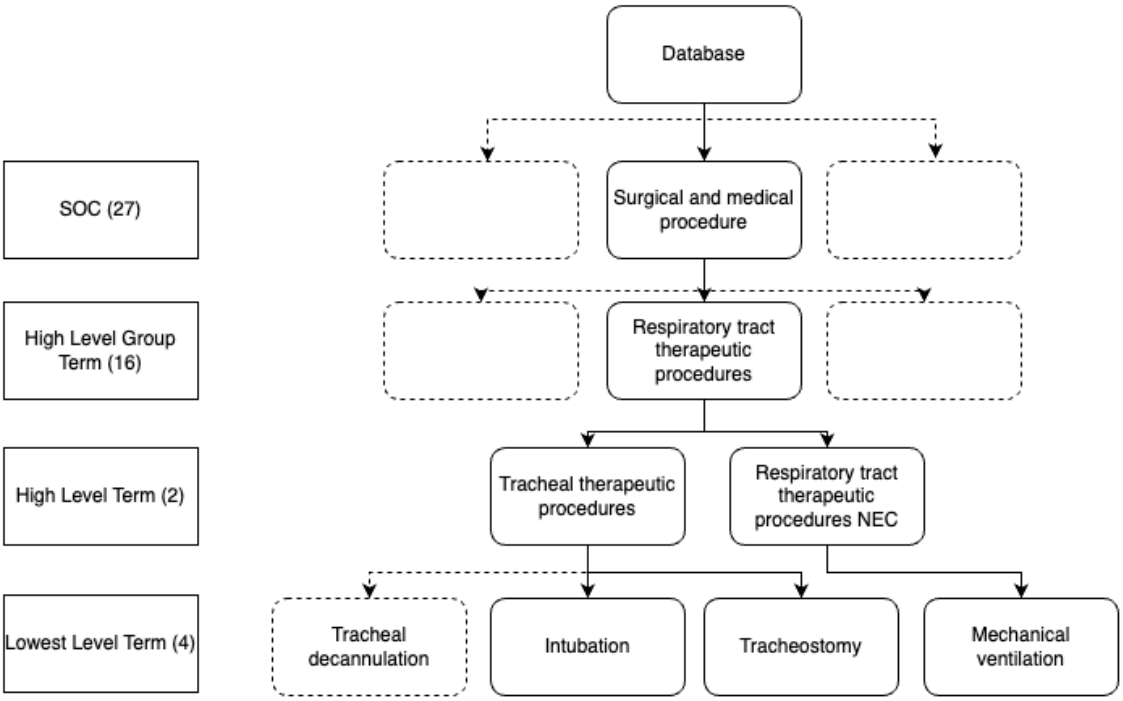


FIGURE D.2: Medical intervention terms related to tracheostomy using Medical Dictionary for Regulatory Activities (MDRA)

Legend: SOC: Standard Of Care, each square represents a level in the medical intervention terms tree, the numbers in parentheses are the numbers of possible terms at each level of the tree, not selected levels and terms are represented with dashed

Appendix E

Carbon footprint

In this appendix we give an order of magnitude of the carbon footprint of this thesis.

Contents

E.1 Introduction

The notion of carbon footprint has gained increasing importance in recent years. It corresponds the total amount of greenhouse gas emissions directly or indirectly attributed to an individual, organization, or product. In this appendix, following the work of (Sauty, 2023), we have assessed the carbon footprint associated with the research conducted for this thesis. The objective was to give insights into our research's contribution to global warming and explore directions for future carbon footprint reduction.

E.2 Method

We estimated the emissions related to the conferences attended during this thesis. Estimation of transportation-related emissions was made using data from the French database ADEME (https://bilans-ges.ademe.fr/), the distances travelled and the mode of transportation. Nevertheless, due to the multimodal transportation from Paris to Thessaloniki, the estimation was made using the website Mollow (`https://www.mollow.eu/destination/ath%C3%A8nes-gr%C3%A8ce`). For conferences abroad, the option by plane was also estimated for comparison.

We also took into account office-related emissions. They were extracted from the work of (Sauty, 2023), and came from an internal committee at the Paris Brain Institute, utilizing the "labo1.5" estimator (`https://labos1point5.org/`).

All the computations were run on individual computers and that specific electricity consumption was thus considered inside the office-related emissions. Transportation from home to office was also neglected as done by bike.

Precise figures were not of interest and the order of magnitude should be kept in mind.

E.3 Results

TABLE E.1: Estimation of the carbon footprint of this thesis.

Legend: For reference, 2000kgCO2-eq/year/person for all emissions (private and professional) is a symbolic threshold believed to be sustainable at the global scale.

		Emissions (kg CO2-eq)	% of emissions
Conferences	Paris - Edinburgh	20	0.23
	Paris - Newcastle	16	0.19
	Paris - Milan	15	0.18
	Paris - Toulouse	12	0.14
	Paris - Thessaloniki	127	1.48
Institute	Electricity	1,000	11.64
	Heating/AC	3,600	41.91
	IT hardware	3,800	44.24
	Total	8,590	100

Compared to the field of deep learning, for which a lot of conferences are outside of Europe, in biostatistics, European conferences of great quality exist which has enabled controlling the carbon footprint associated. For the abroad conferences, the total carbon footprint was 178 kg CO2-eq compared to its equivalent by plane of 2,156 kg CO2-eq. Most of the carbon footprint of this thesis is related to the institute footprint with IT hardware (44.24%) and heating/AC (41.91%).

E.4 Discussion

For reference, 2000kg CO2-eq/year/person for all emissions (private and professional) is a symbolic threshold believed to be sustainable at the global scale. With a carbon footprint of around 8.5 tons, work remains on office-shared facilities.

Efforts were made to reduce the footprint of this thesis using low-carbon transportation. Nevertheless, such a reduction came as an individual responsibility with an associated increased amount of time and money needed. The research community is currently calling for global changes (Tao et al., 2021) and some hybrid conferences have been created to reduce the impact of conferences (https://cuttinggardens2023.org/).

This footprint estimation could be improved by taking into account a more precise footprint of the in-person conferences attended.

E.5 Conclusion

Through this estimation, we aim to promote awareness and advocate for more sustainable practices within academic research, so that it will be less dependent on individual responsibility in the future.

Appendix F

First author journal papers

In this appendix papers published or submitted are available.

Journal papers

1. **J. Ortholand**, P.-F. Pradat, S. Tezenas Du Montcel, S. Durrleman, *Interaction of sex and onset site on the disease trajectory of ALS*, Journal of Neurology, 2023. [hal-04216957]

Submitted journal papers

1. **J. Ortholand**, S. Durrleman, S. Tezenas Du Montcel, *Joint model with latent disease age: overcoming the need for reference time*,[preprint]

Interaction of sex and onset site on the disease trajectory of amyotrophic lateral sclerosis

Juliette Ortholand, Pierre-François Pradat, Sophie Tezenas Du Montcel, Stanley Durrleman

Interaction of sex and onset site on the disease trajectory of Amyotrophic Lateral Sclerosis

Author(s)

Juliette ORTHOLAND[1], Pierre-François PRADAT[2,3,4], Sophie TEZENAS DU MONTCEL[1], Stanley DURRLEMAN[1]

Corresponding Author

Juliette ORTHOLAND, juliette.ortholand@icm-institute.org

Affiliation Information for All Authors

1. Sorbonne Université, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hôpital de la Pitié Salpêtrière, F-75013, Paris, France
2. Sorbonne Université, CNRS, INSERM, Laboratoire d'Imagerie Biomédicale, Paris, France
3. APHP, Département de Neurologie, Hôpital Pitié-Salpêtrière, Centre référent SLA, Paris, France
4. Northern Ireland Centre for Stratified Medicine, Biomedical Sciences Research Institute Ulster University, C-TRIC, Altnagelvin Hospital, Derry/Londonderry, United Kingdom.

Equal Author Contribution

Sophie Tezenas du Montcel and Stanley Durrleman contributed equally to this work as senior authors.

Contributions

- Juliette ORTHOLAND: Drafting/revision of the manuscript for content, including medical writing for content; Study concept or design; Analysis or interpretation of data.
- Pierre-François PRADAT: Drafting/revision of the manuscript for content, including medical writing for content;
- Sophie Tezenas du Montcel: Drafting/revision of the manuscript for content, including medical writing for content; Study concept or design; Analysis or interpretation of data.
- Stanley Durrleman: Drafting/revision of the manuscript for content, including medical writing for content; Study concept or design; Analysis or interpretation of data.

Figure count

3

Table count

1

Search terms

ALS, sex, bulbar, spinal, onset site

Acknowledgement

We warmly thank Igor Koval for the review of the data processing and all the team members of ARAMIS for their input and advice.

Study Funding

This paper is funded in part by grant number 826421 (TVB-Cloud) from the H2020 program, and ANR-10-IAIHU-06 (IHU ICM), ANR-19-P3IA-0001 (PRAIRIE 3IA Institute), and ANR-19-JPW2-000 (E-DADS) from ANR.

Disclosures

The authors report no relevant disclosures.

ABSTRACT

Background: Studies showed the impact of sex and onset site (spinal or bulbar) on disease onset and survival in ALS. However, they mainly result from cross-sectional or survival analysis, and the interaction of sex and onset site on the different proxies of disease trajectory has not been fully investigated.

Methods: We selected all patients with repeated observations in the PRO-ACT database. We divided them into four groups depending on their sex and onset site. We estimated a multivariate disease progression model, named ALS Course Map, to investigate the combined temporal changes of the four sub-scores of the revised ALS Functional Rating Scale (ALSFRSr), the forced vital capacity (FVC), and the body mass index (BMI). We then compared the progression rate, the estimated age at onset, and the relative progression of the outcomes across each group.

Results: We included 1,438 patients from the PRO-ACT database. They were 51% men with spinal onset, 12% men with bulbar onset, 26% women with spinal onset, and 11% women with bulbar onset. We showed a significant influence of both sex and onset site on the ALSFRSr progression. The BMI decreased 8.9 months earlier (95% CI = [3.9, 13.8]) in women than men, after correction for the onset site. Among patients with bulbar onset, FVC was impaired 2.6 months earlier (95% CI = [0.6, 4.6]) in women.

Conclusion: Using a multivariable disease modelling approach, we showed that sex and onset site are important drivers of the progression of motor function, BMI, and FVC decline.

INTRODUCTION

Amyotrophic Lateral Sclerosis (ALS) is an adult-onset heterogeneous neurodegenerative disease that occurs on average between 55 and 60 years. It is characterized by a degeneration of both upper and lower motor neurons[1], and the average survival time from onset to death is three to four years[2]. The disease shows an important phenotypic variability, particularly in terms of the site of disease onset and progression[3]. The difference between the disease starting with dysarthria and dysphagia (bulbar onset) and the disease with limb weakness (spinal onset) is well documented. Among other, they differ in terms of incidence[3]; progression rate[4–6]; and survival duration[4].

Several epidemiological studies have shown a lower incidence of ALS in women compared to men[4,7] as well as differences in the onset site and age at onset[8]. More recently genetic[9] and imaging [10] studies have enriched the knowledge on this sexual dimorphism. However, most of these results come from cross-sectional studies[7,8] or survival[4] studies. By design, the first one failed to capture the longtime effect of clinical or demographic variables, such as age or sex[11]. The second only presents half of the information on the disease progression by studying mortality, leaving aside the function in daily activities. Thus how the marked sexual dimorphism influences the progression and the presentation of each ALS form, has not been fully described.

The originality of the present study is to investigate the interaction of sex and onset site on the whole progression of the following clinical proxies: BMI[12], forced vital capacity (FVC)[13], and sub-scores of ALS functional rating scales (ALSFRSr) [14]. To do so, we have used a recent longitudinal model designed for neurodegenerative diseases[15,16]. This enabled us to study all the outcomes at the same time, consider the non-linearity of the progression of outcomes, and capture patients' heterogeneity better than with classic models.

MATERIALS AND METHODS

Study population

We analyzed patients from the Pooled Resource Open-Access ALS Clinical Trials Consortium (PRO-ACT) database. It is composed of an aggregation of 23 phase II and III clinical trials and one observational study. Information that could lead to any identification, such as the clinical trial, tested drug, study centres, or dates, was not included in the database. As the database is an aggregation, inclusion and exclusion criteria for patients to enter the cohort are multiple and will not be described here, more information could be found in the paper that presents the database [17].

For our study, we selected patients with at least two visits for all the outcomes listed below. We collected: outcome values at all available visits; the sex of the participant; and the onset site (spinal or bulbar), a label that is available in the PRO-ACT database. We divided participants into four groups: men with bulbar onset, women with bulbar onset, men with spinal onset, and women with spinal onset.

Outcomes

We selected a set of outcomes that are important clinical proxies of ALS progression and are available in the PRO-ACT database. We considered the most widely used functional rating system in patients with ALS, namely the revised version of the ALS functional rating scale (ALSFRSr)[14]. The scale starts at a maximum theoretical value of 48 and decreases with the severity of the disease till zero. We considered the four ALSFRSr sub-scores assessing four distinct domains: bulbar, gross motor, fine motor, and respiratory functions. Each sub-scale has 12 points each[14]. In addition, we considered the FCV [13] that we normalized using the computed normal FVC[18] and the BMI[12].

We normalized the outcomes between 0 (the healthiest value) and +1 (the maximum pathological change). The four ALSFRSr sub-scores have been normalized using their theoretical maximum and minimum values. For FVC and BMI, we first applied a box-cox transformation (estimated on the whole PRO-ACT dataset) to get a standard normal distribution[19,20,21]. We then clipped the resulting value between -3 and 3 points and converted them to the 0 to 1 scale.

In addition to these repeated measures, we collected the following covariates when available: riluzole use, trial arm (active, placebo, observational), age at first symptom, age at diagnosis, age at baseline, Mitos[22] and FT9 scores[23] at baseline. These data were not used as input for the disease progression model.

ALS Course Map

We used a disease progression model, named ALS Course Map, which is a multivariate non-linear Bayesian mixed-effects model[15,16]. The model estimates the progression of a series of outcomes from a longitudinal dataset. It estimates a subject-specific set of curves for each outcome and summarizes them at the population, or sub-group level. The model does not require imputation as it is robust to missing data [24]. In addition, it does not require a standardized time between visits [24].

ALS course map models the progression with a logistic pattern of outcomes development. As in other mixed effect models, two types of parameters are computed: fixed effect, referred to as population parameters, and random effect, referred to as individual parameters. Population parameters capture the average ALS progression from which patients deviation could be captured by three individual parameters:

The first individual parameter is the disease time shift, τ_i, expressed in years. It represents the delay between the estimated age at disease onset for patient i and the whole population. For instance, $\tau_i = -1$ means that the estimated disease onset of the patient occurs one year earlier than the average patient in the population. The estimation of the time shift takes all outcomes into account. There is a single time shift for a given subject.

The second one is the progression rate, α_i. It represents a value that multiplies the population progression rate to get the rate of the patient i. For instance, $\alpha_i = 1.2$ means that the disease progresses 1.2 times faster than the average patient in the population. The estimation of the progression rate also takes all outcomes into account. There is a single rate for a given subject.

Finally, a series of inter-marker spacings, ω_{ij}, expressed in months. One ω_{ij} changes the onset of the outcome j of the patient i while keeping other outcomes fixed. For instance, $\omega_{ij} = +2$ means that the j-th outcome changes two months earlier than for the average patient in the population, all other things being equal. Each subject has six inter-marker spacings, one per outcome in the model. Inter-marker spacings are centred (i.e. their weighted sum equals zero) so that they do not interplay with the estimation of the time shift.

We used the open-source software, Leaspy, to estimate the model parameters from a longitudinal dataset. The software is publicly available at: https://gitlab.com/icm-institute/aramislab/leaspy/Experimental design. This disease progression model has been applied to other neurodegenerative conditions[25,26].

Statistical analysis

We trained the model with the whole data set composed of four groups. We assessed the convergence with a rank normalized R-hat diagnostic test using a threshold of 1.2. To study the goodness-of-fit of the model to the actual data, we compared the estimated value of each visit with its real available value to compute the reconstruction error. We compared the reconstruction error among groups.

To test for the interaction of sex and onset site on the ALS course map individual parameters we used ANOVAs. We compared the individual parameters estimated by the ALS Course Map between sex and onset site interaction using ANOVAs. For significant results, we computed an independent pairwise Tukey-Kramer test[27]. Otherwise, we computed an ANOVA without interaction to test the influence of each covariate adjusted for the other. All the results are presented with a 95% confidence interval computed with pairwise Tukey-Kramer or ANOVA without interaction, depending on the interaction significance. We used the same comparison process to describe the cohort and compare the groups at baseline. For the cohort description, results are associated with the corrected sample standard deviation. Statistical tests were performed at a conventional 2-tailed type I error of 0.05. Data were analyzed using python software 3.8, using packages statsmodel[28] and bioinfokit[29].

Standard Protocol Approvals, Registrations, and Patient Consents

Data used in this study came from the PRO-ACT database which is institutional review board–approved. Patients originally consented to participate in clinical trials that have been aggregated in the database. Every patient has been de-identified. We accessed the data following the access procedure detailed here: https://ncri1.partners.org/ProACT/Account/Register.

Data availability

Access to the database PRO-ACT could be asked for through the procedure above. Further information for replication of the study is available in supplementary material in eReproducibility.

RESULTS

Population

Among the 8,571 patients (78,824 visits) in the PRO-ACT database, we dropped 220 values of BMI and 45 values of FVC that were out of the normalization range. We then selected the 1,463 patients with at least two visits for all the outcomes selected. Finally, we selected the 1,438 patients with information about the onset site. Note that sex was available for all patients. Their characteristics are presented in Table 1. The cohort was composed of 51% men with spinal onset, 12% men with bulbar onset, 26% women with spinal onset, and 11% women with bulbar onset. There were no significant differences in follow-up duration (1.00 ± 0.59 years); the number of visits per patient (11.6 ± 7.2 visits); and time between visits (34.5 ± 28.5 days) over the four groups. Nevertheless, we observed significant differences among groups for ages, time durations, and scores at baseline.

ALS Course Map

We can first extract qualitative results from the four groups' trajectories. Figure 1 represents the progression of ALSFRSr subscores for each subgroup. Overall, curves are shifted on the right for patients with bulbar onset (Figure 1 B and D) compared to patients with spinal onset (Figure 1 A and C). Curves are also shifted on the right for women (Figure 1 A and B) compared to men (Figure 1 C and D), illustrating later onset for women compared to men.

The model also exhibited differences in the timing of outcomes. For instance, it confirmed the progression order of ALSFRSr bulbar curve (orange) and ALSFRSr motor curves (red and green): they are switched between patients with spinal onset (Figure 1 A and C) and patients with bulbar onset (Figure 1 B and D).

Finally, we can note differences in speed. Patients with spinal onset (Figure 1 A and C and Figure 2 purple and grey) seem to have steeper curves than patients with bulbar onset (Figure 1 B and D and Figure 2 cyan and pink). All these results are confirmed below by quantitative statistical tests.

Estimated age at disease onset

We found an interaction between the onset site and sex for the estimated disease onset ($p = 0.0020$) (in supplementary materials eFigure 1). For patients with bulbar onset, the disease onset for men occurs 6.0 years earlier (95% CI = [2.6, 9.5]) than for women while for a woman with spinal onset, it occurs 4.7 years earlier (95% CI = [1.9, 7.4]) than for women with bulbar onset.

Rate of progression

We did not find any significant interaction between the onset site and sex for the rate of progression ($p=0.14$) (in supplementary materials eFigure 2). Patients with bulbar onset were found to progress 1.4 times faster (95% CI = [1.2, 1.6]) than patients with spinal onset independently of sex.

Outcomes onset of impairment

Respiratory function

We found an interaction between sex and onset site for FVC ($p<0.001$) but not for ALSFRSr respiratory score ($p=0.058$), once corrected for the estimated age at disease onset and rate of progression (Figure 3 E and F). Among patients with bulbar onset, FVC is impaired 2.6 months earlier (95% CI = [0.6, 4.6]) in women than in men. Among men, FVC is impaired 5.4 months earlier (95% CI = [3.8, 7.0]) in patients with bulbar onset compared to spinal onset. Among women, the difference increases to 8.3 months (95% CI = [6.7, 9.9]). ALSFRSr respiratory scale changes 2.4 months earlier (95% CI = [0.7, 4.0]) for patients with bulbar onset compared to spinal onset independently of the sex.

Motor function

We found an interaction between sex and onset site for gross motor ($p<0.001$) but not for fine motor scores ($p=0.12$), once corrected for the estimated age at disease onset and rate of progression (Figure 3 A and B). Among patients with spinal onset, ALSFRSr gross motor scale deteriorates 4.5 months later (95% CI = [3.1, 6.0]) in women than in men. But ALSFRSr fine motor scale changes 3.5 months earlier (95% CI = [2.6, 4.4]) in women than in men, independently of the onset site. Among men, ALSFRSr gross motor scale changes 10.6 months earlier (95% CI = [8.3, 12.8]) for patients with spinal onset compared to bulbar onset. Among women, the difference increases to 16.0 months earlier (95% CI = [13.7, 18.2]). ALSFRSr fine motor scale changes 10.2 months earlier (95% CI = [9.0, 11.3]) for patients with spinal onset compared to bulbar onset, independently of the sex.

Bulbar function

We did not observe any interaction between sex and onset site for ALSFRSr bulbar score ($p=0.01$), once corrected for the estimated age at disease onset and rate of progression (Figure 3 C). Among patients with bulbar onset, ALSFRSr bulbar scale changes 6.4 months earlier (95% CI = [1.6, 11.1]) in women than men. Among men, ALSFRSr bulbar scale changes 27.2 months later (95% CI = [23.4, 31.0]) for patients with spinal onset compared to bulbar onset. Among women, the difference increases to 32.5 months later (95% CI = [28.7, 36.3]).

BMI

We did not observe an interaction between sex and onset site for BMI ($p=0.17$), once corrected for the estimated age at disease onset and rate of progression (Figure 3 D). BMI decreases 8.9 months earlier (95% CI = [3.9, 13.8]) in women than men, independently of the onset site. BMI decreases 10.0

months earlier (95% CI = [3.8, 16.1]) for patients with bulbar onset compared to spinal onset, independently of the sex.

DISCUSSION

Sex and onset site interact with disease progression in ALS as demonstrated on all the clinical proxies studied, including disability, weight loss, and respiratory function. In particular, several proxies were affected earlier in women than in men.

Concerning the effect on weight, we showed that BMI starts to decrease 9 months earlier in women than men, independently of the onset site. Considering the rapid progression of ALS, with a mean survival of 3-6 years, this inter-sex difference appears clinically meaningful. Weight loss is a classical consequence of bulbar impairment leading to dysphagia. But alterations in body weight are present in ALS patients already decades before clinical manifestation of ALS[30] and are a negative prognostic factor associated with shorter survival [31]. Weight loss is also associated with increased energy expenditure. It is present in almost half of the patients and is associated with shorter survival[32]. Inter-sex differences in metabolism may also affect the disease trajectory. The role of energy and lipid metabolism abnormalities in the pathogenesis of ALS is currently a focus of many studies in animal models and ALS patients[33–36]. Inter-sex differences in the percentage of body fat may also modify disease progression. It is higher in women[37] and may vary with menopausal status[38] . Studies comparing fat mass with impedancemetry or dual-energy X-ray absorptiometry are needed to better investigate the role of metabolism and the effect of sex on disease progression [3939].

The effect of sex on the respiratory decline has not been reported so far. We showed that among patients with bulbar onset, FVC is impaired 2.6 months earlier in women than in men. This difference

is also clinically meaningful since respiratory distress is the main cause of death in ALS and the decline of FVC is the main criterion for the indication of non-invasive ventilation[40].

An unexpected result was that the ALSFRSr fine motor scale changes earlier in women, regardless of the onset site, but that the ALSFRSr gross motor scale changes later in women among patients with spinal onset. Further studies are needed to investigate whether these differences might be related to sexual dimorphism in the anatomic pattern of abnormalities, both in terms of susceptibility to neurodegeneration and compensatory mechanisms. The advances in neuroimaging enable an increasing number of research studies investigating cortical and subcortical structure in both health and disease conditions[41,42]. Note that a dimorphism of regions linked to fine motors has been shown in healthy participants [43]. In addition, significant sex differences in the anatomical patterns of cortical and subcortical pathology have been shown in ALS[41]. Finally, these sex differences might be a bias in the self-evaluation of motor abilities using ALSFRSr.

On the one hand, from a biological perspective, the mechanisms of sexual dimorphism in ALS have been shown at an anatomical level in specific brain regions, in vivo, thanks to imaging[10] and related to sex-specific microglial regulation[44], sex hormones[45], or susceptibility to oxidative stress[46]. A study, performed on a large number of ALS women with natural menopause and well-defined oral contraceptive usage, has demonstrated that longer exposure to female hormones has neuroprotective effects on motoneurons in ALS (positive association between a longer reproductive condition, the susceptibility to ALS, and the survival of ALS patients)[45]. On the other hand, the overall heritability of ALS is estimated to be of approximately 21% using genome-wide association studies (GWAS) datasets[9]. The effect of sex was found to be different according to the type of mutation C9orf72 or SOD1 on ALS progression[9]. Recent studies on presymptomatic ALS phase support the emerging hypothesis of a

general disorder in development, both in animal[47] and humans at cognitive[48] and structural level[49]. From a developmental origin of health and disease perspective[50], the well-known sexual dimorphism in physiological development could also play a role.

A strength of our study is that our model was based on a large, multicentric, and longitudinal cohort. We used a mixed effect model following neurological research recommendation[51]. The model used is robust to missing data and does not require a standardized time between visits [24]. Logistic curves were used to overcome common limitations on score analysis: curvilinearity, and potential floor or ceiling effects[52]. The reconstruction error of the model was of the same range as the one of the Alzheimer's disease study [25], for which the model has been originally developed (eFigure 3 and eFigure 4 in supplementary materials).

Our study has several limitations. PRO-ACT may not be representative of the real-life ALS patient population. Since the database aggregates clinical trial data, there is an overrepresentation of men, young subjects, patients with spinal onset, and slow progressors. Nevertheless, these results could help better understand the progression of placebo arms and design strategies for patient stratification in clinical trials. Due to the high level of death censoring in PRO-ACT, we were not able to analyse the survival in addition to the disease progression. The outcomes selected has also their own limits. BMI does not capture the distinction between fat, muscle, or bone mass. We have also used the sub-scores of ALSFRSr separately as recommended[14], but as lower motor neurons often dominate in the overall clinical presentation [5], the score might be ill-suited for other sub-form of the disease. A population-based study showed the existence of plateaus in ALSFRSr mainly shorter than 6 months, with only 10% of patients with wider plateaus [53]. Thus, modeling the progression of ALSFRSr subscores by logistic curves could be debated. Nevertheless, plateaus appear in high scores and are unlikely to change the

results since our patients' average length of follow-up is one year. The absence of information on genetic mutation[3]; neuropsychological symptoms[3], exposure to metals or agricultural factor[54] and lifestyle[55,56] also limits the interpretation of the results.

To summarize, we showed that sex and onset site interaction is an important driver of ALS progression considering many clinical proxies such as disability, weight loss, and respiratory function. Pathophysiological mechanisms underlying these differences are elusive and may account for an interplay of biological, lifestyle, and environmental factors. Further studies in animal models and patients are needed to better understand their respective contributions.

References

1. Rowland LP, Shneider NA. Amyotrophic Lateral Sclerosis. *N Engl J Med*. 2001;344(22):1688-1700. doi:10.1056/NEJM200105313442207

2. Talbott EO, Malek AM, Lacomis D. The epidemiology of amyotrophic lateral sclerosis. *Handb Clin Neurol*. 2016;138:225-238. doi:10.1016/B978-0-12-802973-2.00013-6

3. Swinnen B, Robberecht W. The phenotypic variability of amyotrophic lateral sclerosis. *Nat Rev Neurol*. 2014;10(11):661-670. doi:10.1038/nrneurol.2014.184

4. Tysnes OB, Vollset SE, Aarli JA. Epidemiology of amyotrophic lateral sclerosis in Hordaland county, western Norway. *Acta Neurol Scand*. 1991;83(5):280-285. doi:10.1111/j.1600-0404.1991.tb04701.x

5. Brown RH, Al-Chalabi A. Amyotrophic Lateral Sclerosis. *N Engl J Med*. 2017;377(2):162-172. doi:10.1056/NEJMra1603471

6. Brooks BR. Natural history of ALS: symptoms, strength, pulmonary function, and disability. *Neurology*. 1996;47(4 Suppl 2):S71-81; discussion S81-82. doi:10.1212/wnl.47.4_suppl_2.71s

7. Logroscino G, Traynor BJ, Hardiman O, et al. Incidence of Amyotrophic Lateral Sclerosis in Europe. *J Neurol Neurosurg Psychiatry*. 2010;81(4):385-390. doi:10.1136/jnnp.2009.183525

8. Chiò A, Moglia C, Canosa A, et al. ALS phenotype is influenced by age, sex, and genetics: A population-based study. *Neurology*. 2020;94(8):e802-e810. doi:10.1212/WNL.0000000000008869

9. F T, G D, S B, G T. Genetics and Sex in the Pathogenesis of Amyotrophic Lateral Sclerosis (ALS): Is There a Link? *Int J Mol Sci*. 2020;21(10). doi:10.3390/ijms21103647

10. Bede P, Elamin M, Byrne S, Hardiman O. Sexual dimorphism in ALS: Exploring gender-specific neuroimaging signatures. *Amyotroph Lateral Scler Front Degener*. 2014;15(3-4):235-243. doi:10.3109/21678421.2013.865749

11. Farrington DP. Longitudinal Research Strategies: Advantages, Problems, and Prospects. *J Am Acad Child Adolesc Psychiatry*. 1991;30(3):369-374. doi:10.1097/00004583-199105000-00003

12. Moglia C, Calvo A, Grassano M, et al. Early weight loss in amyotrophic lateral sclerosis: outcome relevance and clinical correlates in a population-based cohort. *J Neurol Neurosurg Psychiatry*. 2019;90(6):666-673. doi:10.1136/jnnp-2018-319611

13. Daghlas SA, Govindarajan R, Pooled Resource Open-Access ALS Clinical Trials Consortium. Relative effects of forced vital capacity and ALSFRS-R on survival in ALS. *Muscle Nerve*. 2021;64(3):346-351. doi:10.1002/mus.27344

14. Rooney J, Burke T, Vajda A, Heverin M, Hardiman O. What does the ALSFRS-R really measure? A longitudinal and survival analysis of functional dimension subscores in amyotrophic lateral sclerosis. *J Neurol Neurosurg Psychiatry*. 2017;88(5):381-385. doi:10.1136/jnnp-2016-314661

15. Schiratti JB, Allassonnière S, Colliot O, Durrleman S. A Bayesian Mixed-Effects Model to Learn Trajectories of Changes from Repeated Manifold-Valued Observations. Published online 2017:33.

16. Schiratti JB, Allassonniere S, Colliot O, Durrleman S. Learning spatiotemporal trajectories from manifold-valued longitudinal data. Published online 2015:9.

17. Atassi N, Berry J, Shui A, et al. The PRO-ACT database: design, initial analyses, and predictive features. *Neurology*. 2014;83(19):1719-1725. doi:10.1212/WNL.0000000000000951

18. Brändli O, Schindler C, Künzli N, Keller R, Perruchoud AP. Lung function in healthy never smoking adults: reference values and lower limits of normal of a Swiss population. *Thorax*. 1996;51(3):277-283. doi:10.1136/thx.51.3.277

19. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine Learning in Python. *Mach Learn PYTHON*.:6.

20. Yeo I, Johnson RA. A new family of power transformations to improve normality or symmetry. *Biometrika*. 2000;87(4):954-959. doi:10.1093/biomet/87.4.954

21. Box GEP, Cox DR. An Analysis of Transformations. *J R Stat Soc Ser B Methodol*. 1964;26(2):211-252.

22. Fang T, Al Khleifat A, Stahl DR, et al. Comparison of the King's and MiToS staging systems for ALS. *Amyotroph Lateral Scler Front Degener*. 2017;18(3-4):227-232. doi:10.1080/21678421.2016.1265565

23. Thakore NJ, Lapin BR, Kinzy TG, Pioro EP. Deconstructing progression of amyotrophic lateral sclerosis in stages: a Markov modeling approach. *Amyotroph Lateral Scler Front Degener*. 2018;19(7-8):483-494. doi:10.1080/21678421.2018.1484925

24. Couronne R, Vidailhet M, Corvol JC, Lehericy S, Durrleman S. Learning Disease Progression Models With Longitudinal Data and Missing Values. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. IEEE; 2019:1033-1037. doi:10.1109/ISBI.2019.8759198

25. Koval I, Bône A, Louis M, et al. AD Course Map charts Alzheimer's disease progression. *Sci Rep*. 2021;11(1):8020. doi:10.1038/s41598-021-87434-1

26. Couronné R. Progression models for Parkinson's Disease. Published online 2021:194.

27. Tukey JW. Comparing Individual Means in the Analysis of Variance. *Biometrics*. 1949;5(2):99-114. doi:10.2307/3001913

28. Skipper S, Perktold J. statsmodels: Econometric and statistical modeling with python. Published online 2010.

29. Bedre R. reneshbedre/bioinfokit: Bioinformatics data analysis and visualization toolkit. Published online 2020.

30. Peter RS, Rosenbohm A, Dupuis L, et al. Life course body mass index and risk and prognosis of amyotrophic lateral sclerosis: results from the ALS registry Swabia. *Eur J Epidemiol*. 2017;32(10):901-908. doi:10.1007/s10654-017-0318-z

31. Li JY, Sun XH, Cai ZY, et al. Correlation of weight and body composition with disease progression rate in patients with amyotrophic lateral sclerosis. *Sci Rep*. 2022;12(1):13292. doi:10.1038/s41598-022-16229-9

32. Nakamura R, Kurihara M, Ogawa N, et al. Investigation of the prognostic predictive value of serum lipid profiles in amyotrophic lateral sclerosis: roles of sex and hypermetabolism. *Sci Rep*. 2022;12(1):1826. doi:10.1038/s41598-022-05714-w

33. Dupuis L, Pradat PF, Ludolph AC, Loeffler JP. Energy metabolism in amyotrophic lateral sclerosis. *Lancet Neurol*. 2011;10(1):75-82. doi:10.1016/S1474-4422(10)70224-6

34. Nelson AT, Trotti D. Altered Bioenergetics and Metabolic Homeostasis in Amyotrophic Lateral Sclerosis. *Neurother J Am Soc Exp Neurother*. 2022;19(4):1102-1118. doi:10.1007/s13311-022-01262-3

35. Vandoorne T, De Bock K, Van Den Bosch L. Energy metabolism in ALS: an underappreciated

opportunity? *Acta Neuropathol (Berl)*. 2018;135(4):489-509. doi:10.1007/s00401-018-1835-x

36. Guillot SJ, Bolborea M, Dupuis L. Dysregulation of energy homeostasis in amyotrophic lateral sclerosis. *Curr Opin Neurol*. 2021;34(5):773-780. doi:10.1097/WCO.0000000000000982

37. Blaak E. Gender differences in fat metabolism. *Curr Opin Clin Nutr Metab Care*. 2001;4(6):499-502. doi:10.1097/00075197-200111000-00006

38. Toth MJ, Tchernof A, Sites CK, Poehlman ET. Effect of menopausal status on body composition and abdominal fat distribution. *Int J Obes Relat Metab Disord J Int Assoc Study Obes*. 2000;24(2):226-231. doi:10.1038/sj.ijo.0801118

39. Lee I, Kazamel M, McPherson T, et al. Fat mass loss correlates with faster disease progression in amyotrophic lateral sclerosis patients: Exploring the utility of dual-energy x-ray absorptiometry in a prospective study. *PloS One*. 2021;16(5):e0251087. doi:10.1371/journal.pone.0251087

40. Morelot-Panzini C, Bruneteau G, Gonzalez-Bermejo J. NIV in amyotrophic lateral sclerosis: The 'when' and 'how' of the matter. *Respirology*. 2019;24(6):521-530. doi:10.1111/resp.13525

41. Bede P, Murad A, Lope J, Hardiman O, Chang KM. Clusters of anatomical disease-burden patterns in ALS: a data-driven approach confirms radiological subtypes. *J Neurol*. 2022;269(8):4404-4413. doi:10.1007/s00415-022-11081-3

42. Trojsi F, Di Nardo F, Caiazzo G, et al. Between-sex variability of resting state functional brain networks in amyotrophic lateral sclerosis (ALS). *J Neural Transm Vienna Austria 1996*. 2021;128(12):1881-1897. doi:10.1007/s00702-021-02413-0

43. Lissek S, Hausmann M, Knossalla F, et al. Sex differences in cortical and subcortical recruitment during simple and complex motor control: An fMRI study. *NeuroImage*. 2007;37(3):912-

926. doi:10.1016/j.neuroimage.2007.05.037

44. Lay K, Li G. Do microglial sex differences contribute to sex differences in neurodegenerative diseases? *Trends Mol Med*. 2019;25(9):741-749. doi:10.1016/j.molmed.2019.05.001

45. Raymond J, Mehta P, Larson T, Pioro EP, Horton DK. Reproductive History and Age of Onset for Women Diagnosed with Amyotrophic Lateral Sclerosis: Data from the National ALS Registry: 2010-2018. *Neuroepidemiology*. 2021;55(5):416-424. doi:10.1159/000516344

46. Sumien N, Cunningham JT, Davis DL, et al. Neurodegenerative Disease: Roles for Sex, Hormones, and Oxidative Stress. *Endocrinology*. 2021;162(11):bqab185. doi:10.1210/endocr/bqab185

47. Martin E, Cazenave W, Allain AE, Cattaert D, Branchereau P. Implication of 5-HT in the Dysregulation of Chloride Homeostasis in Prenatal Spinal Motoneurons from the G93A Mouse Model of Amyotrophic Lateral Sclerosis. *Int J Mol Sci*. 2020;21(3):1107. doi:10.3390/ijms21031107

48. Lulé DE, Müller HP, Finsel J, et al. Deficits in verbal fluency in presymptomatic *C9orf72* mutation gene carriers—a developmental disorder. *J Neurol Neurosurg Psychiatry*. 2020;91(11):1195-1200. doi:10.1136/jnnp-2020-323671

49. Bertrand A, Wen J, Rinaldi D, et al. Early Cognitive, Structural, and Microstructural Changes in Presymptomatic C9orf72 Carriers Younger Than 40 Years. *JAMA Neurol*. 2018;75(2):236-245. doi:10.1001/jamaneurol.2017.4266

50. Stephens MC, Brandt V, Botas J. The developmental roots of neurodegeneration. *Neuron*. 2022;110(1):1-3. doi:10.1016/j.neuron.2021.12.004

51. Locascio JJ, Atri A. An Overview of Longitudinal Data Analysis Methods for Neurological Research. *Dement Geriatr Cogn Disord Extra*. 2011;1(1):330-357. doi:10.1159/000330228

52. Gordon PH, Cheng B, Salachas F, et al. Progression in ALS is not linear but is curvilinear. *J Neurol*. 2010;257(10):1713-1717. doi:10.1007/s00415-010-5609-1

53. Vasta R, D'Ovidio F, Canosa A, et al. Plateaus in amyotrophic lateral sclerosis progression: results from a population-based cohort. *Eur J Neurol*. 2020;27(8):1397-1404. doi:10.1111/ene.14287

54. Borghero G, Sechi MM, Vasta R, et al. Spatial clustering of amyotrophic lateral sclerosis in Sardinia, Italy: The contribution of age, sex, and genetic factors. *Muscle Nerve*. 2023;n/a(n/a). doi:10.1002/mus.27939

55. Trojsi F, Sagnelli A, Vanacore N, et al. Clinical features and lifestyle of patients with amyotrophic lateral sclerosis in Campania: brief overview of an Italian database. *Ann Ist Super Sanita*. 2012;48(3):287-291. doi:10.4415/ANN_12_03_09

56. Westeneng HJ, van Veenhuijzen K, van der Spek RA, et al. Associations between lifestyle and amyotrophic lateral sclerosis stratified by C9orf72 genotype: a longitudinal, population-based, case-control study. *Lancet Neurol*. 2021;20(5):373-384. doi:10.1016/S1474-4422(21)00042-9

Table 1: Patient characteristics Legend: data are expressed as mean± standard deviation (percentage of available data) or frequency (percentage), p-value: ANOVA interaction p-value, BM:

Outcomes			Spinal Man	Bulbar Man	Spinal Woman	Bulbar Woman	p-value	Subgroup relations
Numbers	Patients		757	126	405	150		
	Visits		9236	1447	4771	1799		
	Visits per patient		12.2 ± 10.0	11.5 ± 7.4	11.8 ± 9.0	12.0 ± 7.7	0.46	-
Riluzole use	Took at least once		533 (70%)	82 (65%)	286 (71%)	96 (64%)		
	Never took		126 (17%)	31 (25%)	72 (18%)	29 (19%)		
	No information on use		98 (13%)	13 (10%)	47 (12%)	25 (17%)		
Trial arm	Active		378 (50%)	66 (52%)	204 (50%)	65 (43%)		
	Placebo		210 (28%)	35 (28%)	115 (28%)	40 (27%)		
	Observational		169 (22%)	25 (20%)	86 (21%)	45 (30%)		
Age	First symptoms		52.3 ± 11.1 (100%)	53.1 ± 11.6 (100%)	54.1 ± 10.9 (100%)	59.5 ± 9.5 (100%)	0.0021	(SM,BM,SW)<BW
	Diagnosis		53.0 ± 11.1 (87%)	53.7 ± 11.7 (89%)	54.7 ± 10.6 (85%)	60.5 ± 9.5 (90%)	0.0012	(SM,BM,SW)<BW
	Baseline		54.1 ± 11.1 (100%)	54.6 ± 11.7 (100%)	56.0 ± 10.8 (100%)	61.1 ± 9.4 (100%)	0.0018	(SM,BM,SW)<BW
Time duration	Patient time of follow-up (y)		1.0 ± 0.7 (100%)	1.0 ± 0.6 (100%)	1.0 ± 0.8 (100%)	1.0 ± 0.6 (100%)	0.58	-
	Time between visits (d)		34 ± 28 (100 %)	35 ± 28 (100 %)	35 ± 30 (100 %)	34 ± 29 (100 %)	0.33	-
	Disease duration (y)		1.8 ± 1.0 (100%)	1.4 ± 0.9 (100%)	1.9 ± 1.1 (100%)	1.6 ± 0.9 (100%)	0.61	B<S***
Outcomes at baseline	Mitos		0.2 ± 0.5 (59%)	0.2 ± 0.5 (59%)	0.3 ± 0.5 (61%)	0.2 ± 0.5 (71%)	0.70	-
	FT9		1.8 ± 0.9 (59%)	1.9 ± 1.0 (59%)	1.9 ± 0.9 (61%)	1.9 ± 1.0 (71%)	0.52	-
	ALSFRSr score	Total	37.2 ± 5.4 (59%)	37.4 ± 5.8 (59%)	35.9 ± 5.5 (61%)	37.3 ± 5.3 (71%)	0.19	W<M***
		Respiratory	11.3 ± 1.3 (59%)	11.1 ± 1.3 (59%)	11.2 ± 1.3 (61%)	11.1 ± 1.2 (71%)	0.73	-
		Bulbar	10.8 ± 1.5 (59%)	7.6 ± 2.1 (59%)	10.5 ± 1.7 (61%)	6.7 ± 2.3 (71%)	0.025	BW<BM<(SW,SM)
		Fine motor	7.6 ± 2.8 (59%)	9.2 ± 2.6 (59%)	8.0 ± 2.7 (61%)	10.0 ± 2.1 (71%)	0.39	M<W**;B<S***
		Gross motor	7.6 ± 2.5 (59%)	9.5 ± 2.5 (59%)	6.2 ± 2.4 (61%)	9.5 ± 2.3 (71%)	0.0012	SW<SM<(BW,BM)
	Forced vital capacity		75.9 ± 14.4 (97%)	70.3 ± 13.6 (96%)	74.6 ± 14.4 (96%)	68.7 ± 13.8 (97%)	0.83	B<S***
	BMI		27.1 ± 4.2 (98%)	26.8 ± 4.3 (97%)	26.4 ± 5.0 (98%)	25.6 ± 4.9 (98%)	0.46	W<M***

bulbar men, BW: bulbar women, SM: spinal men, SW: spinal women, *: <0.05, **: <0.01, *: <0.001**

Figure 1 Average progression of ALSFRSr subscores depending on sex and onset site: Respiratory(blue), Bulbar (orange), Fine Motor (red), Gross Motor (green)

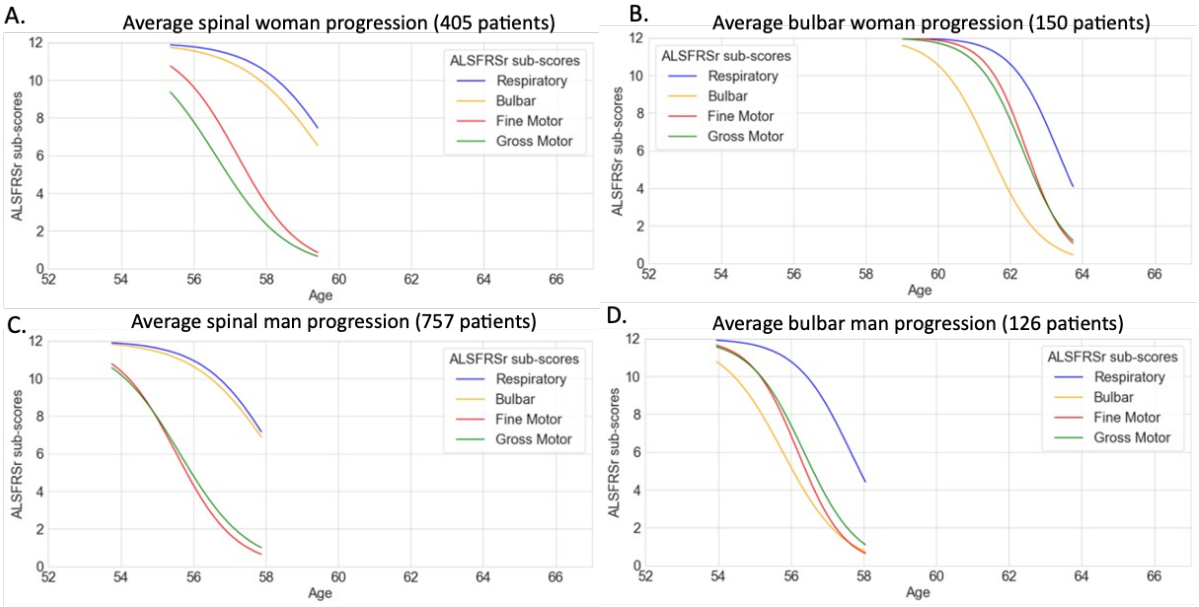
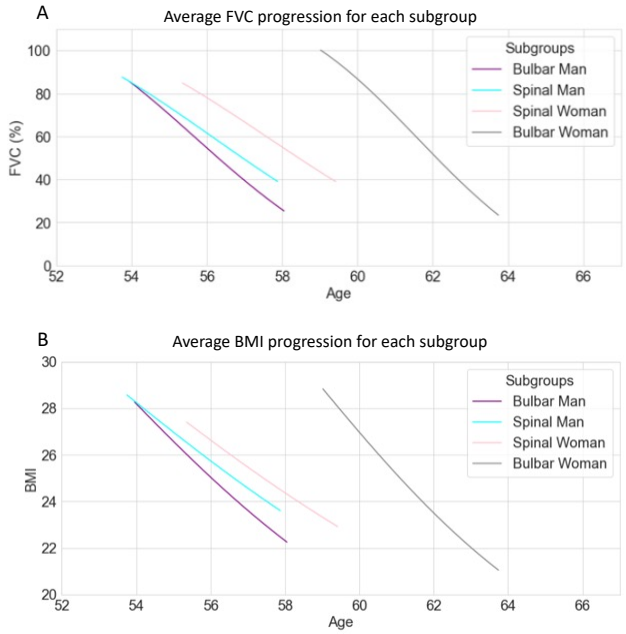


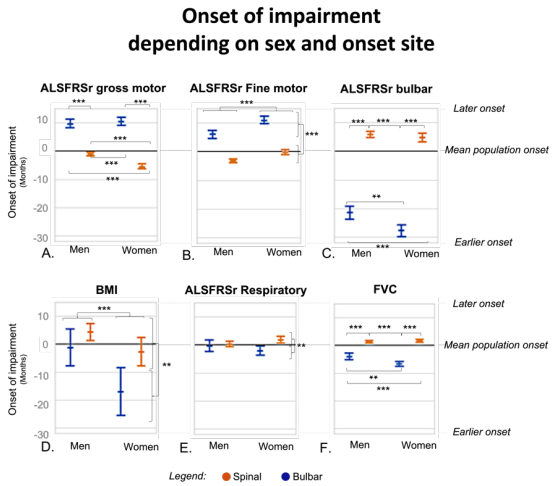
Figure 2: Average subgroup progressions of FVC and BMI for each subgroup



Legend: FVC: Forced vital capacity, BMI: Body mass index

Figure **3**: **Onset of impairment depending on sex and onset site**

The vertical axis presents the delay of outcome impairment onset in month compared to the mean onset of the whole population. The horizontal axis is for sex with estimation for the bulbar patients in blue and the spinal patients in orange. The graph displays the mean of this delay with its confidence interval. Results are corrected for the group rate of progression and estimated age at onset. ANOVA interaction p-value is below 0.001 for ALSFRSr gross motor, ALSFRSr bulbar, and FVC and equal to 0.058 for ALSFRSr respiratory, 0.12 for ALSFRSr fine motor, and 0.17 for BMI. FVC: Forced Vital Capacity (%), **: p-value for pairwise comparison <0.01 ,*: p-value for pairwise comparison <0.001**



Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

Joint model with latent disease age: overcoming the need for reference time

Juliette ORTHOLAND* | Nicolas GENSOLLEN | Stanley DURRLEMAN | Sophie TEZENAS DU MONTCEL

[1]ARAMIS, Sorbonne Universite, Institut du Cerveau - Paris Brain Institute - ICM, CNRS, Inria, Inserm, AP-HP, Hopital de la Pitie Salpetriere, Paris, France

Correspondence

*Juliette ORTHOLAND Email: juliette.ortholand@inria.fr

Abstract

Introduction Heterogeneity of the progression of neurodegenerative diseases is one of the main challenges faced in developing effective therapies. With the increasing number of large clinical databases, disease progression models have led to a better understanding of this heterogeneity. Nevertheless, these diseases may have no clear onset and biological underlying processes may start before the first symptoms. Such an ill-defined disease reference time is an issue for current joint models, which have proven their effectiveness by combining longitudinal and survival data.

Objective In this work, we propose a joint non-linear mixed effect model with a latent disease age, to overcome this need for a precise reference time.

Method To do so, we utilized an existing longitudinal model with a latent disease age as a longitudinal sub-model and associated it with a survival sub-model that estimates a Weibull distribution from the latent disease age. We then validated our model on different simulated scenarios. Finally, we benchmarked our model with a state-of-the-art joint model and reference survival and longitudinal models on simulated and real data in the context of Amyotrophic Lateral Sclerosis (ALS).

Results On real data, our model got significantly better results than the state-of-the-art joint model for absolute bias (4.21(4.41) versus 4.24(4.14)(p-value=1.4e-17)), and mean cumulative AUC for right censored events (0.67(0.07) versus 0.61(0.09)(p-value=1.7e-03)).

Conclusion We showed that our approach is better suited than the state-of-the-art in the context where the reference time is not reliable. This work opens up the perspective to design predictive and personalized therapeutic strategies.

1 | INTRODUCTION

Neurodegenerative disorders are an important burden for the healthcare system[1]. The heterogeneity of the progression of these diseases is a major challenge in the development of effective therapies, as in Parkinson's Disease (PD)[2], Alzheimer's Disease (AD)[3] or Amyotrophic Lateral Sclerosis (ALS)[4]. With the increasing number of large clinical databases, the development of disease progression models has helped in understanding this heterogeneity better. Two main data types can be subjected to modelling: longitudinal data such as repeated measures of clinical scores or biomarkers; or survival data, with the occurrences of events like death, surgical intervention or entry into an institution. In some circumstances, the U.S. Food and Drug Administration might demand assessing treatment efficacy using both types of outcomes, as for ALS[5].

Data available for neurodegenerative disease is often sparse and covers only parts of the progression, which emphasises the need to realign different patients' ages depending on their disease stages to extract a full typical timeline of the disease[6]. The date of the first symptom is often used to do so, even if it does not capture the difference in terms of speed of progression. But in such disease the underlying biological processes start earlier than the first specific symptoms of the disease, making the first symptoms reported by the patient not representative of the disease onset and very subjective. For instance, in Alzheimer's Disease, the progression of amyloid and neurodegeneration biomarkers starts before the manifestation of the first clinical symptoms[7]. Similarly in ALS, changes in metabolism start before a significant weight loss and the first motor symptoms[8], which again complicates the establishment of reliable reference time points for monitoring disease progression. Thus two challenges are faced to extract a full typical progression of these diseases: mapping the disease onset and the speed of progression of the patients.

Survival and longitudinal data are often associated with the same biological processes. In such a case, modelling both data together results in more precise estimates and improves inference[9]. Joint models enable us to do so. They are composed of three parts: a model for the longitudinal data, a model for the survival data and a linkage, often a shared latent variable, which captures the association between the two types of data[10,11,12,13,14].
Classical survival models are hazard-based regression models[15]. The most used is the Cox Proportional Hazard (Cox) model[16], with an effect on the hazard scale. Its main interests are that it does not require estimating the baseline hazard function when used alone and that it is easy to interpret due to the proportional impact of covariates on the baseline hazard. Nevertheless, the baseline hazard needs to be estimated for joint models[10] and the proportional assumption is often violated on long follow-up. Another family of models is the Accelerated Failure Time (AFT) model[17]. In these models, covariates directly affect survival time, but modelling the hazard function is mandatory, making it less used for survival analysis alone. Besides, the results may be harder to interpret.
Classical longitudinal models are mixed-effect models that allow for modelling repeated and correlated observations[18]. To capture the heterogeneity of the patient progression, in addition to population parameters, named fixed effects, they have individual parameters, named random effects. These random effects capture the individual variability around the average disease progression. Linear mixed-effects models (LMM) are the simplest ones, yet they are often inadequate as changes in clinical outcomes are non-linear over long periods of time, due to curvilinearity and potential floor or ceiling effects[19]. The most common type of Non-Linear Mixed effects are Generalised Linear Mixed-effect Models (GLMMs), a direct extension of LMM, which enables the estimation of non-linear progression[20]. Such classical mixed-effect models rely on an empirical disease time axis which limits their temporal resolution to the resolution of the reference time used to index the disease time axis[21,22,6]. To overcome this issue, longitudinal models that capture the data-driven disease timeline behind the observable data, named data-driven disease progression models, have been developed[6]. Among others, disease Progression score models have

been developed[23] and extended[24,25] as well as models that allow direct integration of knowledge about different stages of diseases[26]. Finally, some models have explored the creation of a latent disease age[27,28,21,22]. A non-linear-mixed effect model based on Riemannian geometry was proposed to capture a latent disease age, from a population sample, without the need for a reference time[21,22]. This model showed good performances in degenerative disease by outperforming the 56 alternative methods for predicting cognitive decline in Alzheimer's disease in the TADPOLE challenge[29 30].

For the linkage structure between survival and longitudinal data, two types of modelling have been proposed: the latent class model[31 32] and the shared random effects model[33]. Such an approach aggregates similar patients which could help better understand the heterogeneity even though the meaning of the different classes remains unknown. It allows to calculate the probability of an individual belonging to a particular class but may result in some patients being almost equitably distributed. Shared random effect models are often used to avoid these limits. Nevertheless, they have their own limitation: by including predictors of the longitudinal outcome in the survival model, they usually focus on how longitudinal outcomes impact survival. Whatever their type, all these joint models rely on GLMMs, necessitating the use of reliable reference times, which is not available in our context.

In this paper, we propose a latent age joint model suited for neurodegenerative diseases that overcomes the need for a reliable reference time of the state-of-the-art joint models. To do so, we used an existing longitudinal model with a latent disease age[21,22] as the longitudinal sub-model and used its defined latent disease age as the linkage structure. We associated a survival sub-model that estimates a Weibull distribution from the latent disease age. After introducing the proposed joint model, we validate it using a sensitivity analysis on simulated data corresponding to different real-like clinical scenarios. We then benchmark the proposed joint model against reference models on simulated and real ALS data, and show that the proposed approach is better suited in the context of the absence of a reliable reference time point reference time than the state-of-the-art.

2 | MODEL

2.1 | Generic Framework

2.1.1 | Data

We consider N patients, associated with longitudinal data: repeated measures of one given outcome y. Each patient i is followed for n_i visits. For each visit j, we denote $t_{i,j}$ the age at which the outcome is measured, and $y_{i,j}$ the value of the outcome for the patient at this visit.

We assume that we also observe an event e, and denote t_{e_i} the age of the patient when the event is observed. Nevertheless, the event may occur after the follow-up period. In this case, the event is said to be censored, in opposition to observed. To distinguish censored and observed events, a boolean B_{e_i} is associated with the time of the event t_{e_i}: $B_{e_i} = 0$ if the event is censored and $B_{e_i} = 1$ if the event is observed. If the event is censored, the time of the last visit is used as the time of the event[34].

2.1.2 | Joint model structure

The objective of joint models is to describe the combination of two types of clinical data: longitudinal data and survival data, with their relationship. Here, the longitudinal process, $\gamma_i(t)$, is the progression of an outcome, measured by $y_{i,j}$ at each time $t_{i,j}$ for each visit j of the patient i. The longitudinal process is estimated with a Gaussian noise $\epsilon \sim \mathcal{N}(0, \sigma)$ compared to the measure, so that: $y(t_{i,j}) = \gamma(t_{i,j}) + \epsilon_{i,j}$. The survival process $S_i(t)$, the probability that a patient i experiences the event after age t ($S_i(t) = p(t_{e_i} > t)$).

To describe the model, we will further use the formalism of mixed-effects models[18]. Such models are composed

of two types of parameters: parameters that differ from one patient to the other and enable to encapsulate the individual variability, named random effects, and parameters that capture the population specificity and are shared by all the patients, named fixed effects.

2.2 | The Proposed Joint model

A non-linear mixed-effect model with a latent disease age was first introduced by Schiratti et al.[21] and was more broadly used for longitudinal process modelling of neurodegenerative diseases[22,35]. Both the latent age and the modelling of the longitudinal process presented below are extracted from Schiratti et al.[21,22].

2.2.1 | The latent disease age: correction of individual variation

The idea of the latent disease age $\psi_i(t)$ is to map the chronological age of a patient into a latent disease age t_d representative of the disease stage of the patient. Using the formalism described before, it can be written as :

$$\psi_i(t) = e^{\xi_i}(t - \tau_i) + t_0 \tag{1}$$

where e^{ξ_i} is the progression rate of patient i, τ_i is its time-shift and t_0 is the population estimated reference time.

In the proposed joint model, the idea is to encapsulate all the individual variability of the patient i, in the latent disease age ψ_i, with random effects. The latent disease age is then used as the link between the longitudinal and survival processes $(\gamma_i(t), S_i(t))$, which are estimated from the latent disease age with the composition of functions that describe only the population (γ_0, S_0) and are shared by all patients with fixed effects.

$$\begin{cases} \gamma_i(t) = \gamma_0(\psi_i(t_{i,j})) \\ S_i(t) = S_0(\psi_i(t)) \end{cases}$$

For the survival modelling, instead of using time 0 as a start time , we use the reference time t_0 and impose $\forall t < t_0, S_0(t) = 1$. Indeed, t_0 is estimated thanks to the visit times and corresponds to a time of a given value of the score that most of the patients experimented with. Thus at that time, patients should not be dead. We have checked that even for the existing longitudinal model most patients were still alive at t_0.
Compared to joint models with shared random effects models of Rizopoulos et al.[33], where $S_0(\psi_i(t)) = f(\gamma_0(\psi_i(t))$ with f a function of fixed parameters, the Proposed model only depends on the latent disease age, that could be seen as an undergoing process.

2.2.2 | Modelling longitudinal process

The modelling of the longitudinal process consists in computing the trajectory from the latent disease age defined in part 2.2.1. We will study a clinical score with curvilinearity, and potential floor or ceiling effects[19]. Thus a logistic function will be used to model the outcome value from the latent disease age t_d. It is parametrized as follows:

$$\gamma_0(t_d) = \left(1 + g \times \exp(-v_0 \frac{(g+1)^2}{g}(t_d - t_0))\right)^{-1} \tag{2}$$

where t_0 is the population estimated reference time defined in 2.2.1, v_0 is the speed of the logistic curve at t_0 and $\frac{1}{1+g}$ is the value of the logistic curve at t_0. To get the real value of the outcome $y_{i,j}$, the latent disease age $\psi_i(t)$ is first applied, then the longitudinal process from the latent disease age $\gamma_0(t)$ and finally a Gaussian noise $\epsilon_{i,j}$ is added. Note that we assume here that all the noises of visits are independent. The whole longitudinal process can thus be written as:

$$y_{i,j} = \gamma_0(\psi_i(t_{i,j})) + \epsilon_{i,j} = \gamma_i(t_{i,j}) + \epsilon_{i,j} \tag{3}$$

2.2.3 | Modelling survival process

The Weibull distribution is used to model the survival probability from the latent disease age t_d:

$$S_0(t_d) = \mathbb{1}_{t_d \geq t_0} \exp\left(-\left(\frac{t_d - t_0}{\nu}\right)^{\rho}\right) + \mathbb{1}_{t_d < t_0}$$

where ν represents the variability of the distribution and ρ the shape of the data distribution. To get the individual survival probability, we apply first the latent disease age and then the survival structure:

$$\begin{aligned} S_i(t) = & \quad S_0(\psi_i(t)) \\ = & \ \mathbb{1}_{\psi_i(t) > t_0} \exp\left(-\left(\frac{\psi_i(t) - t_0}{\nu}\right)^{\rho}\right) + \mathbb{1}_{\psi_i(t) \leq t_0} \\ = & \quad \mathbb{1}_{t > \tau_i} \exp\left(-\left(\frac{e^{\xi_i}(t - \tau_i)}{\nu}\right)^{\rho}\right) + \mathbb{1}_{t \ \leq \tau_i} \end{aligned}$$

From there we also compute the individual hazard, which is, assuming that a patient has survived for a time t, the probability that he will not survive for an additional time dt:

$$\begin{aligned} h_i(t) = & \quad -\frac{S_i'(t)}{S_i(t)} \\ = & \ \mathbb{1}_{\psi_i(t) > t_0} \frac{\rho e^{\xi_i}}{\nu} \left(\frac{\psi_i(t) - t_0}{\nu}\right)^{\rho - 1} \\ = & \quad \mathbb{1}_{t > \tau_i} \frac{\rho e^{\xi_i}}{\nu} \left(\frac{e^{\xi_i}(t - \tau_i)}{\nu}\right)^{\rho - 1} \end{aligned}$$

2.2.4 | Proposed joint model

The proposed joint model referred to after as the Proposed model, is thus the combination of both a longitudinal sub-model $\gamma_i(t)$ and a survival sub-model $S_i(t)$ using the latent disease age $\psi_i(t)$ as a linkage structure:

$$\begin{cases} \psi_i(t) = e^{\xi_i}(t - \tau_i) + t_0 \\ \gamma_i(t) = \gamma_0(\psi_i(t)) = \left(1 + g \times \exp(-\upsilon_0 \frac{(g+1)^2}{g}(\psi_i(t) - t_0))\right)^{-1} \\ S_i(t) = S_0(\psi_i(t)) = \mathbb{1}_{\psi_i(t) > t_0} \exp\left(-\left(\frac{\psi_i(t) - t_0}{\nu}\right)^{\rho}\right) + \mathbb{1}_{\psi_i(t) \leq t_0} \end{cases} \tag{4}$$

2.3 | Estimation

2.3.1 | Parameters

We consider that all parameters of the structural model, described in part 2.2.4, are latent parameters following a given distribution. From now on, we make a distinction between fixed and random effects. Random effects follow a Gaussian distribution. Most fixed effects are assumed to be positive. To do this, they are not directly estimated, but their logarithm is estimated. The new variable is written with a tilde in relation to the original variable (for instance, $\tilde{\upsilon}_0 = \log(\upsilon_0)$). Log fixed effects follow a Gaussian distribution with the average and the standard deviation defined as hyperparameters. To ensure identifiability, we set $\overline{\xi} = 0$ and $t_0 = \overline{\tau}$. We end up with the following structure:

- Latent parameters (z):
  - Latent fixed effects (z_{fe}): log fixed effects sampled from Gaussian distributions parametrized by model parameters and hyperparameters $z_{fe} = \{\tilde{g}, \tilde{\upsilon}_0, \tilde{\nu}, \tilde{\rho}\}$,
  - Latent random effects (z_{re}): estimated from Gaussian distributions parametrized by model parameters $z_{re} = \{\xi_i, \tau_i\}$,

- Model parameters (θ): fixed effects sampled from log-likelihood maximisation $\theta = \{\sigma_\xi, \sigma_\tau, t_0, \overline{\xi}, \overline{\tau_0}, \overline{\nu}, \overline{\rho}, \sigma\}$
- Hyperparameters (Π): set by the user $\Pi = \{\sigma_{\tilde{g}}, \sigma_{\tilde{\nu_0}}, \sigma_{\tilde{\nu}}, \sigma_{\tilde{\rho}}\}$

2.3.2 | Log-likelihood

Log-likelihood can be divided into two different terms: data attachment which represents how well the model describes the data (y, t_e, B_e) and a prior attachment, which prevents over-fitting.

$$\log p((y, t_e, B_e), z, \theta \mid \Pi) = \log p(y, t_e, B_e \mid z, \theta, \Pi) + \log p(z, \theta \mid \Pi)$$

The first term, data attachment, can be divided again into two terms considering that survival and longitudinal processes are independent regarding random effects. This is a quite common assumption in other papers [10 31]. We can also separate the prior attachment term: two terms for the prior attachment of latent parameters (fixed and random) and one term for the prior attachment of model parameters. We end up with the following expression :

$$\begin{aligned}\log p((y, t_e, B_e), z, \theta \mid \Pi) = &\quad \log p(y \mid z, \theta, \Pi) \\ &+ \log p(t_e, B_e \mid z, \theta, \Pi) \\ &+ \quad \log p(z_{fe} \mid \theta, \Pi) \\ &+ \quad \log p(z_{re} \mid \theta, \Pi) \\ &+ \qquad \log p(\theta \mid \Pi)\end{aligned}$$

The different log-likelihood parts are described below associated with their different assumptions. Note that the total formula of the log-likelihood is available in annexe A.

Longitudinal data attachment

For longitudinal process modelling, we assume that patients and their visits are independent and that the noise of the process follows a Gaussian distribution. We thus get:

$$\log p(y \mid z, \theta, \Pi) = \sum_{i,j} \log p(y_{i,j} \mid z, \theta, \Pi) = \sum_{i,j} -\log\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2\sigma^2}\left(y_{i,j} - \gamma_0\left(\psi_i\left(t_{i,j}\right)\right)\right)^2$$

Survival data attachment

For survival process modelling, we once again assume that all patients are independent and that the modelling of the survival process depends on whether the event is observed or not:

$$\begin{aligned}\log p(t_e, B_e \mid z, \theta, \Pi) = &\quad \textstyle\sum_i \log p(t_{e_i}, B_{e_i} \mid z, \theta, \Pi) \\ = &\ \textstyle\sum_i \mathbb{1}_{B_{e_i}} \times \log\left(h_i(t_{e_i})\right) + \sum_i \log\left(S_i(t_{e_i})\right)\end{aligned}$$

Note that if $\psi_i(t) < t_0$ $\log\left(h_i(t_{e_i})\right) = -\infty$. To prevent estimation issues, we initialise the algorithm at a possible point getting inspiration from barrier methods.

Latent random effects priors attachment

As patients are supposed to be independent of each other, we suppose that random effects are independent conditionally to θ and Π. The regularization term associated, with $\overline{\xi} = 0$, is then:

$$\begin{aligned}\log p(z_{re} \mid \theta, \Pi) = &\quad \log p(\tau_i \mid \theta, \Pi) + \log p(\xi_i \mid \theta, \Pi) \\ = &\ -N \log\left(\sigma_\tau\sqrt{2\pi}\right) - \tfrac{1}{2\sigma_\tau^2}\textstyle\sum_i(\tau_i - t_0)^2 \\ &-N \log\left(\sigma_\xi\sqrt{2\pi}\right) - \tfrac{1}{2\sigma_\xi^2}\textstyle\sum_i(\xi_i - \overline{\xi})^2\end{aligned}$$

Latent fixed effects prior attachment

Each latent fixed effect is independently sampled from a posterior distribution. The regularization term associated is then:

$$\begin{aligned}
\log p(z_{lat} \mid \theta, \Pi) = \; & \log p(\tilde{g} \mid \theta, \Pi) + \log p(\tilde{v_0} \mid \theta, \Pi) \\
+ \; & \log p(\tilde{v} \mid \theta, \Pi) + \log p(\tilde{\rho} \mid \theta, \Pi) \\
= \; & -\log\left(\sigma_{\tilde{g}}\sqrt{2\pi}\right) - \tfrac{1}{2\sigma_{\tilde{g}}^2}\left(\tilde{g} - \overline{\tilde{g}}\right)^2 \\
- \; & \log\left(\sigma_{\tilde{v}_0}\sqrt{2\pi}\right) - \tfrac{1}{2\sigma_{\tilde{v}_0}^2}\left(\tilde{v}_0 - \overline{\tilde{v}_0}\right)^2 \\
- \; & \log\left(\sigma_{\tilde{v}}\sqrt{2\pi}\right) - \tfrac{1}{2\sigma_{\tilde{v}}^2}\left(\tilde{v} - \overline{\tilde{v}}\right)^2 \\
- \; & \log\left(\sigma_{\tilde{\rho}}\sqrt{2\pi}\right) - \tfrac{1}{2\sigma_{\tilde{\rho}}^2}\left(\tilde{\rho} - \overline{\tilde{\rho}}\right)^2
\end{aligned}$$

Model parameter prior attachment

Priors regarding model parameters are supposed to incorporate knowledge about the disease in the model. As we suppose that the user has no prior knowledge, we turned these priors into non-informative ones, by artificially setting the standard deviation of each to infinity.

2.3.3 | Estimation Algorithm

The first step is calibration, it enables us to estimate fixed and associated random effects from a training data set. Directly maximising the log-likelihood has no analytical solution. Thus we use an Expectation-Maximization algorithm. Nevertheless, the computation of the expectation is also intractable due to the nonlinearity of the model. Thus, we use a Monte-Carlo Markov Chain Stochastic Approximation Expectation-Maximization (MCMC-SAEM) algorithm, as for the existing Longitudinal model. Its convergence has been proven by Kuhn et al.[36] for models that lie in the curved exponential family. The Proposed model falls into such a category and further details are given in annexe B. To get the mean of the distribution of the model, we apply a Robbins-Monro convergence algorithm to the last iterations[37]. More details are given by Koval et al.[35] (p.41-43) and Schiratti et al.[22] (p.106). Latent parameters (defined in 2.3.1), are estimated during the estimating phase of the EM algorithm and model parameters during the maximisation phase, using sufficient statistics. The total log-likelihood, the sufficient statistics and the maximisation update rules, necessary for the computation, are given in the annexe A, B C.

The second step is personalisation. It is used to compute the random effects for new patients from a test data set. During this step, we use previously computed fixed effects from calibration. Thus, only the random effects are estimated, to get their longitudinal and survival processes. The solver *minimise* from the package Scipy[38] was used to maximise the log-likelihood. Note that for predictions, the survival probability is then corrected using the survival probability at the last visit as in other packages[33].

An implementation of the Proposed model is available in the open-source library leaspy (v2): https://gitlab.com/icm-institute/aramislab/leaspy.

2.3.4 | Attenuation estimation bias

Attenuation bias is a well-known phenomenon caused by noise in the input data. In the case of a basic linear regression $Y = aX + b$, the fact that we only have access to $\hat{X} = X + \epsilon$ with ϵ a random noise, leads to an underestimation of the absolute value of the factor parameter a[39]. In the Proposed model, we estimate the latent disease age $\hat{\psi}_i(t)$ for each patient, but we never have access to the exact latent disease age $\psi_i(t)$. Such uncertainty

could lead to an underestimation of two fixed effects ν^{-1} and $\upsilon_0 \frac{(1+g)^2}{g}$ in both the survival and longitudinal process.

$$S_0(\hat{\psi}_i(t)) = \mathbb{1}_{\hat{\psi}_i(t)>t_0} \exp\left(-\left(\frac{\hat{\psi}_i(t)-t_0}{\nu}\right)^{\rho}\right) + \mathbb{1}_{\hat{\psi}_i(t)\leq t_0}$$
$$\gamma_0(\hat{\psi}_i(t)) = \left(1 + g \times \exp(-\upsilon_0 \frac{(1+g)^2}{g}(\hat{\psi}_i(t) - t_0))\right)^{-1}$$

As reported by Frost et al.[40], attenuation bias is not an issue for prediction and should only be dealt with when interpreting the fixed effects of the model. Several options exist to correct them[41,42,43] but none was well suited to our problem. Indeed, we cannot estimate the error on the latent disease age. Thus, we decided to correct our fixed effects using the mean of random effects of the training sample computed during personalisation. This method will be evaluated to make sure that it enables us to be closer to the simulated data.

2.4 | Reference models

2.4.1 | Reference models

We chose to benchmark the Proposed model against several reference models. First, we use one-process-only models. For the survival model, we use a Weibull AFT model to describe the survival process, using the Lifelines package[44]. This model will be referred to as the AFT model. For the longitudinal model, we use the existing Longitudinal model described in part 2.2 using the open-source leaspy library https://gitlab.com/icm-institute/aramislab/leaspy. This model will be referred to as the Longitudinal model. We expect the Proposed model to be at least as good as these two models.

Second, we use a two-stage model: a survival model that uses the random effects of the Longitudinal model as covariates[45]. Even though this model is subject to an immortal bias, it enables us to compare our model to a better survival model than the AFT model. We use the Longitudinal model to extract random effects for each individual, and then use them as covariates in a Weibull AFT model, using the Lifelines package[44]. This model will be referred to as the Two-stage model, and the Proposed model is expected to be at least as good as it.

Third, we use a joint model with shared random effects, to evaluate if the newly proposed structure could improve estimation. To do so, we use a logistic longitudinal process, using the JMbayes2 package[33]. This model will be referred to as the JMbayes2 model. All the model structure equations are summarised in annexe D in Table D1.

2.4.2 | Clinically meaningful estimated parameters

Predictions of different models can easily be compared. Nevertheless, joint models are also used to make cohort descriptions. Since each model has its own parameters, their comparison is difficult. To overcome this issue, the idea is to extract clinically meaningful estimated parameters that can be easily used in cohort description. For the survival process, we choose the median survival time, referred to as the median. For the longitudinal process, we first assume a logistic shape of the curve. We then use the axis value and the slope at the inflexion point as clinically meaningful estimated parameters. We later refer to them as midpoint and growth. Note that for all the different models, we reported the formula used in table D2 in the annexe.

3 | SIMULATION: ROBUSTNESS TO STUDY DESIGN AND PERFORMANCE BENCHMARK

3.1 | Data

3.1.1 | Data simulation method

Data were simulated under the Proposed model structure. We used the following procedure:

1. We simulated random effects using $\xi_i \sim \mathcal{N}\left(0, \sigma_\xi^2\right)$ and $\tau_i \sim \mathcal{N}\left(t_0, \sigma_\tau^2\right)$.
2. We modelled the age at first visit $t_{i,0}$ as $t_{i,0} = \tau_i + \delta_{f_i}$ with $\delta_{f_i} \sim \mathcal{N}\left(\overline{\delta_f}, \sigma_{\delta_f}^2\right)$.
3. We set a time of follow-up per patient T_{f_i}, with $T_{f_i} \sim \mathcal{N}\left(\overline{T_f}, \sigma_{T_f}^2\right)$ and a time between two visits $\delta_{v_{i,j}} = t_{i,j-1} - t_{i,j}$, with $\delta_{v_i,j} \sim \mathcal{N}\left(\overline{\delta_v}, \sigma_{\delta_v}^2\right)$ to simulate n_i visits until $t_{n_i} \leq t_{i,0} + T_{f_i} < t_{n_{i+1}}$.
4. We set the value of the outcome at each visit using $y_{i,j} = \gamma_0(\psi_i(t_{i,j})) + \epsilon_{i,j}$ with $\epsilon_{i,j} \sim \mathcal{N}\left(0, \sigma^2\right)$.
5. For each patient, we simulated the event T_{e_i} using $T_{e_i} \sim e^{-\xi_i}\mathcal{W}(\nu, \rho) + \tau_i$.
6. We considered that the event stopped the follow-up and that the follow-up censored the event. Thus all the visits after the event were censored: $t_{i,j} > T_{e_i}$ and events after the last visit were censored: $t_{i,max(j)} < T_{e_i}$.

Due to random assignment, some patients experienced the event (death) before their first visit. This censure also changed the random effects distribution, as most of the patients censored were progressing fast. To consider this bias, we corrected the true values of clinically meaningful estimated parameters, using the available true value of random effects.

3.1.2 | Simulated Scenarios

We simulated ALS real-like data with different data collection designs (noise, number of visits and patients). We used an ALS dataset, PRO-ACT, described in part 4.1, to get estimated real-like values for parameters. Parameters directly associated with the disease have been extracted from data analysis, using the Longitudinal and AFT models (Figures E1, E2). For each scenario that studies the impact of a given parameter, we simulated three different datasets, that corresponded to an easy, medium and complex situation.

Our first scenario concerned the number of patients needed for calibration. Due to PRO-ACT properties, we decided to test three sample sizes: N = 200, 500, and 1,000, with other parameters being equal to real-like ones. The second scenario enabled us to study the impact of the time of follow-up. Patients in PRO-ACT are followed for an average of 0.96 years. We have tested three different average follow-up durations $\overline{T_f} = 0.75, 1., 1.25$. Visit frequency impact was studied with our third scenario. In PRO-ACT, patients are seen on average every 1.47 months. We have tested three different average times between visits $\overline{\delta_v} = 3, 1.5, 1$. Our fourth scenario enabled us to measure the impact of longitudinal noise in data collection. On PRO-ACT, we have found a noise of 5% on the longitudinal outcome using the Longitudinal model. Note that the test-retest error of ALSFRSr is around 2%[46]. We have thus decided to test $\sigma = 5\%$, 10% and 20%. Finally, our fifth scenario concerned noise in event distribution. We have found a ρ on the survival process of 2.25. We have decided to test three events distribution $(\rho, \nu) = (3, 3.11)$,(2.25, 3.62), (2, 3.84) so that the mode of the distribution $(\nu\left(\frac{\rho-1}{\rho}\right)^{\frac{1}{\rho}})$ does not change but the variation increases.

To make the results of our scenarios comparable, we put the same real-like dataset for all the medium difficulties of the scenarios except for the patient and longitudinal noise one, for which the real-like dataset corresponds to the easier one. We thus ended up with 11 distinct datasets for different study designs, all inspired by real data for which simulation parameters are summarised in the table F3 in the annexe.

3.2 | Method

3.2.1 | Sensitivity analysis

To make our sensitivity analysis on all the simulated scenarios (Table F3 in annexe), we first trained the Longitudinal model during 2,000 iterations. Then, we ran the Proposed model with 70,000 iterations (on average an hour) using the values of the Longitudinal model as initialisation. We then evaluated the quality of estimation in two steps.

Effects estimation

We wanted to assess how well the model could estimate fixed and random effects from simulated data whatever the study design. For fixed effects, as done by Pan et al.[47], we extracted data from the posterior distribution between the 30.000 and 60.000 iterations (just before the Robbins-Monro convergence phase[37]). The performance was assessed using bias compared to true fixed effects that were used for simulation. To reduce the computation complexity of random effects, we decided to extract the mean of the same range of iterations for each individual. As a metric, we used the intraclass correlation between the mean of each individual and the true value that enabled the simulation.

Attenuation bias correction

As described in method 2.3.4, the Proposed model might be subjected to attenuation bias. We proposed a method to reduce this bias and evaluate it, whatever the study design. To do so, clinically meaningful estimated parameters, that are more interpretable than fixed effects, were used. First, the mean of fixed effects was computed through the Robbins-Monro convergence (iteration 70,000). Then we ran personalisation on the training data and corrected the extracted mean with the average of random effects. We use these corrected fixed effects to compute corrected clinically meaningful estimated parameters. In parallel, clinically meaningful estimated parameters were computed using the posterior distribution. We studied if the corrected value was in the 95% confidence interval (CI) of the clinically meaningful estimated parameters using the posterior distribution.
The impact of the correction method on random effects was also evaluated. We used the results of the personalisation for random effects and computed the intraclass correlation between the personalised random effects and the true value that enabled the simulation. Results were compared with the one from part 3.2.1 experiments.

3.2.2 | Benchmark experiment on simulated data

All estimations and predictions of the models were compared using the real-like dataset (Table F3 in annexe). We selected patients with at least 3 visits to ease the following prediction setup. A 10-fold cross-validation (90% - 10%) was used, with 70,000 iterations for each model (on average half an hour), initialised with parameters of the Longitudinal model ran with 2,000 iterations.

Clinically meaningful estimated parameters on simulated data

As we could not compare the fixed effects of the different models, we chose to compare the clinically meaningful estimated parameters, with the true one that enabled simulation.
For each model, we computed clinically meaningful estimated parameters (Table D2 in annexe) and applied the dilution bias correction for both the Proposed model and the Longitudinal model described in part 3.2.1. Then, the bias to true clinically meaningful estimated parameters was computed. We compared the 10-fold pooled bias distribution of the Proposed model with the 4 reference models for the 3 clinically meaningful estimated parameters using a Wilcoxon signed-rank test with a Bonferroni adjustment for multiple pairwise comparisons.

Prediction on simulated data

Finally, we wanted to evaluate both the goodness of survival and longitudinal predictions of the Proposed model against reference models on simulated data. The fixed effects computed through the Robbins-Monro convergence

(iteration 70,000) were used as a predictive model and will be referred to as the prediction model. Personalisation was made using the first two visits of new patients and predictions were estimated on the remaining. The goodness of longitudinal predictions was assessed using absolute errors. The goodness of survival predictions was assessed using the C-index, for event order at 1 and 1.5 years, as in the ALS challenge[48]. Nevertheless, this method is not proper for evaluation[49], we thus added the mean cumulative dynamic AUC at 1 and 1.5 years for a correct measure, but kept the C - index for comparison with existing results. We complement the metrics with Integrated Brier Score (IBS), for absolute distance between real and estimated events. The predictions were compared using a Wilcoxon signed-rank test with a Bonferroni adjustment for multiple pairwise comparisons. All the survival metrics were computed using the package sksurv[50].

3.3 | Results

3.3.1 | Data simulated

The number of visits simulated was relatively stable to 8,341 (31) except for the patient, visits and follow-up scenarios (Figure G3). The censuring rate was between 80 and 85% for all datasets, except for the easy and hard dataset of the follow-up scenario, with 60% and 90% censure (Figure G3).

3.3.2 | Sensitivity analysis

Effects estimation

We evaluated the posterior distribution of fixed effects (Figure H4 in annexe). As expected with the attenuation bias, ν^{-1} was underestimated. ν was overestimated for the real-like dataset by 8.7 (2.9) % and reached the worst results for the harder dataset of the follow-up scenario (Figure H4 in annexe). g and $\upsilon 0$ were also impacted but shared the attenuation bias effect which results in smaller biases. The total bias on the growth will be described after. For the real-like dataset, g was underestimated by -5.2 (3.9) % and υ_0 by -6.7 (2.8)%. The worst results were reached for the harder dataset of longitudinal noise scenarios (Figure H4 in annexe). A less expected result was that ρ was overestimated for the real-like dataset by 30.2 (5.5) % and often across scenarios (Figure H4 in annexe). This could also be linked to the attenuation bias. For the four remaining parameters ($\sigma, \bar{\tau}, \sigma_\tau, \sigma_\xi$), average biases were below 4% for the real-like dataset. Worst results were reached for the harder dataset of longitudinal noise scenarios ($\sigma = 20\%$) for all these fixed effects (Figure H4 in annexe). The intra-class correlation between simulated random effects and the mean of posterior random effects for each patient was between 0.75 and 0.97(Table H4 in annexe).

Attenuation bias correction

The attenuation correction bias was evaluated on clinically meaningful estimated parameters. Almost all corrected clinically meaningful estimated parameters had a lower bias compared to those of the posterior, whatever the scenario and difficulty of the dataset (Figure 1).

The corrected median absolute bias was below 0.43 years for a real value of 4.48 (0.12) years, for all scenarios except the harder of patients and follow-up scenarios. The worst result was reached for the harder follow-up scenario ($\overline{T_f} = 0.75$). The corrected growth absolute bias was below 0.76 points for a real value of -13.19 (0.21) points of ALSFRSr per year, for all scenarios, except for the harder longitudinal noise dataset ($\sigma = 20\%$). The corrected absolute bias midpoint was lower than 0.13 years for a real value of 2.92 (0.06) years, whatever the scenario. The worst result was reached for the harder dataset of the longitudinal noise scenario ($\sigma = 20\%$). All the detailed values for biases are available in Table I5 in the annexe.

Random effects computed through personalisation were not different from those of the posterior mean (Figure I5 in annexe). Detailed values can be found in Table I6 in annexe.

3.3.3 | Benchmark experiment on simulated data

Clinically meaningful estimated parameters of reference models on simulated data

For the longitudinal process, The Proposed and Longitudinal models were closer to the Truth model compared to JMbayes2 (Figure J6 A). Over 10-fold cross-validation, the midpoint was simulated to be 2.97 (0.02) years and the growth of -12.97 (0.11) points of ALSFRSr per year. The Proposed model had a smaller absolute bias on both the growth and midpoint compared to the JMbayes2 model: 0.59 (0.08) compared to 1.71 (0.12) points of ALSFRSr per year (p-value = 3.9e-03) and 0.06 (0.01) compared to 0.43 (0.04) years (p-value = 3.9e-03)(Table 1). This was expected due to the structure of the simulated data. The Longitudinal model had a smaller absolute bias on both the growth and the midpoint compared to the Proposed model: 0.39 (0.08) compared to 0.59 (0.08) (p-value =3.9e-03) and 0.04 (0.01) compared to 0.06 (0.01) (Table 1). This might be because the Longitudinal model has fewer constraints and is applied to an easy task: the simulated data and the Longitudinal model have the same structure.
For the survival process. The Two-stages model and the Proposed model were close to the Truth at the beginning of the progression. The JMbayes2 model overestimates the average patient survival (Figure J6 B). It was confirmed by an absolute bias of 2.34 (0.14) years on the median compared to a truth value of 4.57 (0.03) years over 10-fold cross-validation (Table 1). This issue might be linked to the non-linearity of the random effects in the simulation process. The computed bias on the median was not significantly different for the Proposed model and the Two-stages model. The computed bias on the median was lower for the Proposed model than for the AFT model: respectively 0.09 (0.04) and 0.8 (0.08) (p-value = 5.9e-03) (Table 1).

Prediction on simulated data

On simulated data, 6,382 longitudinal predictions were made at 0.58 (0.41) years from the last visit.
For the longitudinal process, the Proposed model had a significantly smaller longitudinal absolute prediction bias compared to the JMbayes2 model: respectively 3.74 (3.8) and 4.09 (4.1) points of ALSFRSr (p-value = 3.9e-27)(Table 2). The Longitudinal model had a smaller absolute prediction bias compared to the Proposed model: respectively 3.73 (3.82) compared to 3.74 (3.8) (p-value = 3.3e-05).
For the survival process, the Proposed model significantly improved all metrics compared to the JMbayes2 and AFT models. Again, for the JMbayes2 model, this might be due to the failure of the JMbayes2 model to handle the non-linear relation between random effects. The survival predictions of the Proposed model were not significantly distinct for any metric from the one of the Two-stages model.

3.4 | Intermediate conclusion

We have evaluated the impact of attenuation bias on the Proposed model fixed effects and proposed a correction that enabled us to reduce the absolute bias on clinically meaningful estimated parameters. On simulated data, the Proposed model got results close to the Longitudinal model. Results in the survival process were in the same range as the ones of the Two-stage model and outperformed the AFT model. All these results were promising and enabled us to validate the Proposed model on simulated data.

4 | APPLICATION TO ALS DATA: VARIATION IN COHORT DESCRIPTION AND PREDICTION PERFORMANCE BENCHMARK

4.1 | ALS Data: PRO-ACT

Study population

We applied our method to data from an extraction of 2022 of the Pooled Resource Open-Access ALS Clinical Trials Consortium (PRO-ACT) database. It is composed of an aggregation of 23 phase II and III clinical trials and one observational study. It is a pseudonymised set of data with multiple inclusion and exclusion criteria for patients to enter the cohort[51]. For the study, we selected patients with at least three visits for the longitudinal outcomes described below (to ease prediction setup), age at first symptoms, symptom onset (spinal or bulbar) and sex.

Outcomes

As described by the FDA[5], for ALS, clinical trials must demonstrate a treatment effect on function in daily activities and death. Following this guideline, we considered the most widely used functional rating system in patients with ALS, namely the revised version of the ALS functional rating scale revised (ALSFRSr) as longitudinal outcome[52]. The scale starts at a maximum theoretical value of 48 and decreases with the severity of the disease till zero. For computation reasons, it was normalized using its scale so that the 0 value was the healthiest score and +1 the maximum disease score change. Tracheotomy was associated with death as a survival outcome, as in many ALS studies and challenges[48].

4.1.1 | Method - Benchmark experiment on real data

We compared the estimation and prediction of the reference models on real data, using a 10-fold cross-validation (90% - 10%). The Proposed model was run with 70,000 iterations for each model (on average an hour and a half) and initialised with parameters of the Longitudinal model ran for 2,000 iterations.

Clinically meaningful estimated parameters on real data

We computed clinically meaningful estimated parameters from the Proposed model and the reference models to see how different they could be.

Prediction on real data

We wanted to evaluate both the goodness of survival and longitudinal predictions of the Proposed model against reference models, this time on real data. We used the same method as the one developed in part 3.2.2, but applied to real ALS data.

4.2 | Results

4.2.1 | Data description

PRO-ACT datasets had similar or easier characteristics for estimation, compared to our real-like simulated dataset, in terms of the number of patients, visits and time of follow-up (Table 3).

4.2.2 | Benchmark on real data

Clinically meaningful estimated parameters on real data

For the longitudinal process. The Proposed model and Longitudinal model were close compared to the JMbayes2 model (Figure J7 A). Growth in points of ALSFRSr per year ranged from -13.95 (0.08) for the Longitudinal model

to -15.351 (0.1) for the JMbayes2 model, thus a difference of almost 1.4 points per year. The midpoint ranged from 2.78 (0.01) for the Proposed model to 3.22 (0.014) for the JMbayes2 model, thus a difference of more than 5 months (Table 4).
Survival curves were pretty close, even though the variation over 10-fold cross-validation was wider for the JMbayes2 model (Figure J7 B). Apart from the AFT model, the computed medians were close for all the models, ranging from 4.14 (0.02) for the Two-stages model to 4.47 (0.09) for the JMbayes2 model thus a difference of almost 4 months (Table 4).

Prediction on real data

On real data, 18,077 longitudinal predictions were made at 0.63 (0.55) years from the last visit.
For the longitudinal process, the Proposed model had a significantly smaller absolute prediction bias compared to JMbayes2 respectively 4.21 (4.41) and 4.24 (4.14) points of ALSFRSr (p-value = 1.4e-17), but a larger absolute prediction bias compared to the Longitudinal model (4.18 (4.38) (p-value = 2.2e-73)) (Table 5).
For the survival process, the Proposed model was significantly better than all the other models for ordering events, with a mean AUC of 0.67 (0.07) (Table 5). The distance to the observed failure time was not significantly different from the one of the JMbayes2 model with an IBS of 0.1 (0.01), but it was significantly smaller than the one of the Two-stage model (0.11 (0.01) (p-value = 2.5e-04)) and the one of the AFT model (0.12 (0.01) (p-value = 1.1e-04)) (Table 5).

5 | DISCUSSION

Our work showed the potential of a latent disease age joint model to overcome the need for a precise reference time in neurodegenerative disease. After validating the Proposed model on simulated data, we showed how it could improve prediction to forecast ALS progression. Compared to the AFT model and the Two-stage model, the Proposed model outperformed significantly all the metrics, but the Longitudinal model performed slightly better than the Proposed model. Event censoring may not bring much to the Longitudinal model in this case, however, joint modelling corrected shortcomings of the two-stage model, such as immortal bias. Compared to the JMbayes2 model, the Proposed model outperformed the longitudinal and event ordering metrics (C-index and AUC). This shows that our latent disease age might be better suited to capture the heterogeneity of the progression of degenerative diseases. Note that for the event distance metric (IBS) the JMbayes2 model and the Proposed model did not perform significantly differently. This might be because the survival function of the JMbayes2 model, was more flexible, using splines instead of a Weibull function. Finally, we compared the C-index performances of the Proposed model, with those of the ALS challenges that also used PRO-ACT data[48]. With a C-Index around 0.7 (0.5), we ranked seventh, even though we used only ALSFRSr progression to help the prediction, the other models being deep-learning models with multiple covariates. Longitudinal prediction got an absolute bias of 4.18 (4.38) points of ALSFRSr which is correct for a scale design on 48 points but that could still be improved compared to the Minimum Detectable Change (MDC) of 1.59 points of ALSFRSr[53].

Prediction performances can be assessed in real datasets, using hidden visits. The Longitudinal model gave the best longitudinal predictions. The clinically meaningful estimated parameters were close between the Longitudinal and the Proposed models. The difference between the joint models on real data was 1.18 points of ALSFRSr per year on the slope and more than 5 months on the midpoint. These differences were wider than the absolute bias observed in the real-like simulated data. For survival, the Proposed model outperformed the other models on the real data set. Compared to the JMbayes2 model, we observe a difference of almost 4 months for the

median survival. This difference was of the same magnitude as the median absolute bias observed on the real-like simulated data (Table 1).

We have tested the Proposed model in different clinical trial scenarios to give future users insights into biases on clinically meaningful estimated parameters and guidelines as recommended [12]. Note that we studied ALS disease with its own pathological and clinical study characteristics and that guidelines should be adapted in other contexts. We showed that the Proposed model tends to overestimate the survival median and underestimate absolute growth with scenarios increasing difficulty (fewer points, larger noise). Still, the following recommendations could help keep this bias reasonably small. The parameter that impacted the estimation precision the most, was the noise on longitudinal data. In clinical scores, it could be assessed with the Minimal Detectable Change (MDC) and can vary from one disease to another: 10 % for MDS-UPDRS III (Parkinson), [54], 10 % MMSE (Alzheimer) [55] and 3% ALSFRSr (ALS) [53]. In our simulation study, a longitudinal noise of 20% was too large to get good results, but good results were reached with 10 % in our setting (corrected bias: [median: 0.07 year, growth: -0.59 point of ALSFRSr/year, midpoint: -0.02 year]). The number of patients and the time of follow-up should also be considered. For ALS, with a 4 to 5 years median of death from first symptoms, 6 months of follow-up induces too much bias on the estimated median. Results with 1 year of follow-up, that correspond to real-like studies, were good (corrected bias: [median: 0.005 year, growth: 0.45 point of ALSFRSr/year, midpoint: -0.06 year]). Note that observation time in disease timing might also play a role here but we have not further investigated it. We showed that even with only 200 patients, the Proposed model has reasonable biases on clinically meaningful estimated parameters (corrected bias: [median: 0.58 year, growth: 0.37 point of ALSFRSr/year, midpoint: 0.01 year]). Note that there might be a balance with the number of visits (8 by default here). For 1,000 patients, we could reduce the number of visits to 4 per patient every 3 months again with a reasonable bias (corrected bias: [median: 0.14 year, growth: 0.43 point of ALSFRSr/year, midpoint: -0.07 year]). Note that the visits were well-distributed during the whole follow-up time.

The main strength of our model, compared to the state-of-the-art is that it alleviates the proportional hazard hypothesis while keeping an interpretable structure with the latent disease age. Compared to the JMbayes2 model, an underlying shared process is captured by the random effects and not a longitudinal-specific trend which offers a modelling alternative.
Future work could be done to alleviate the assumptions about the independence of longitudinal noises at each visit, which has already been addressed in some models [31]. The Proposed model has also been shown to be subjected to attenuation bias, and although we attempted to address it, we still need to consolidate the hypothesis and the estimation of a good correction for parameters like the ρ parameter. To continue developing this model, future improvements could include adding more flexibility to the survival function, integrating covariates and modelling several longitudinal outcomes or events.

In conclusion, the proposed joint model with latent disease age enabled us to improve the performance of most prediction metrics compared to existing joint models and alleviate the need for a precise reference time. This model opens up the perspective to design predictive and personalized therapeutic strategies.

6 | ACKNOWLEDGEMENTS

Data used in the preparation of this article were obtained from the Pooled Resource Open-Access ALS Clinical Trials (PRO-ACT) Database. As such, the following organizations and individuals within the PRO-ACT Consortium contributed to the design and implementation of the PRO-ACT Database and/or provided data, but did not participate in the analysis of the data or the writing of this report: ALS Therapy Alliance, Cytokinetics, Inc., Amylyx Pharmaceuticals, Inc., Knopp Biosciences, Neuraltus Pharmaceuticals, Inc., Neurological Clinical Research

Institute, MGH, Northeast ALS Consortium, Novartis, Prize4Life Israel, Regeneron Pharmaceuticals, Inc., Sanofi, Teva Pharmaceutical Industries, Ltd., The ALS Association.

7 | FUNDING INFORMATION

The research leading to these results has received funding from the program "Investissements d'avenir" by the French government under management of Agence Nationale de la Recherche, reference ANR-10-IAIHU-06, ANR-19-P3IA-0001 (PRAIRIE 3IA Institute), ANR-19-JPW2-000 (E-DADS) and by H2020 programme reference 826421 (TVB-Cloud).

8 | DATA AVAILABILITY STATEMENT

The Proposed model implementation is available on GitLab on the open-source package leaspy (v2): https://gitlab.com/icm-institute/aramislab/leaspy. The code for data simulation and experiments is available on a public gitlab repository: https://gitlab.com/JulietteOrtholand/jm_article. Note that to run all the experiments, access to a private library leaspype (on demand) and to the PRO-ACT dataset will be required (https://ncri1.partners.org/ProACT/Document/DisplayLatest/2#anchor_12).

References

1. Zahra W, Rai SN, Birla H, et al. The Global Economic Impact of Neurodegenerative Diseases: Opportunities and Challenges. In: Keswani C., ed. *Bioeconomy for Sustainable Development*Singapore: Springer. 2020 (pp. 333–345)

2. Greenland JC, Williams-Gray CH, Barker RA. The clinical heterogeneity of Parkinson's disease and its therapeutic implications. *European Journal of Neuroscience* 2019; 49(3): 328–338. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ejn.14094doi: 10.1111/ejn.14094

3. Duara R, Barker W. Heterogeneity in Alzheimer's Disease Diagnosis and Progression Rates: Implications for Therapeutic Trials. *Neurotherapeutics* 2022; 19(1): 8–25. doi: 10.1007/s13311-022-01185-z

4. Beghi E, Mennini T, Bendotti C, et al. The heterogeneity of amyotrophic lateral sclerosis: a possible explanation of treatment failure. *Current Medicinal Chemistry* 2007; 14(30): 3185–3200. doi: 10.2174/092986707782793862

5. FDA/CDER/"Bastings E. Amyotrophic Lateral Sclerosis: Developing Drugs for Treatment Guidance for Industry. 2019: 11.

6. Young AL, Oxtoby NP, Garbarino S, et al. Data-driven modelling of neurodegenerative disease progression: thinking outside the black box. *Nature Reviews Neuroscience* 2024: 1–20. Publisher: Nature Publishing Groupdoi: 10.1038/s41583-023-00779-6

7. Jack CR, Knopman DS, Jagust WJ, et al. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *The Lancet. Neurology* 2010; 9(1): 119–128. doi: 10.1016/S1474-4422(09)70299-6

8. Peter RS, Rosenbohm A, Dupuis L, et al. Life course body mass index and risk and prognosis of amyotrophic lateral sclerosis: results from the ALS registry Swabia. *European Journal of Epidemiology* 2017; 32(10): 901–908. doi: 10.1007/s10654-017-0318-z

9. Lu X, Chekouo T, Shen H, Leon dAR. A two-level copula joint model for joint analysis of longitudinal and competing risks data. *Statistics in Medicine* 2023; 42(12): 1909–1930. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.9704doi: 10.1002/sim.9704

10. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. New York: Chapman and Hall/CRC . 2012

11. Elashoff RM, Li G, Li N. *Joint Modeling of Longitudinal and Time-to-event Data*. CRC Press, Taylor & Francis Group . 2016. Google-Books-ID: 0JevcQAACAAJ.

12. Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Medical Research Methodology* 2016; 16(1): 117. doi: 10.1186/s12874-016-0212-5

13. Furgal AK, Sen A, Taylor JM. Review and Comparison of Computational Approaches for Joint Longitudinal and Time-to-Event Models. *International Statistical Review* 2019; 87(2): 393–418. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/insr.12322doi: 10.1111/insr.12322

14. Alsefri M, Sudell M, García-Fiñana M, Kolamunnage-Dona R. Bayesian joint modelling of longitudinal and time to event data: a methodological review. *BMC Medical Research Methodology* 2020; 20(1): 94. doi: 10.1186/s12874-020-00976-2

15. Rubio FJ, Remontet L, Jewell NP, Belot A. On a general structure for hazard-based regression models: An application to population-based cancer research. *Statistical Methods in Medical Research* 2019; 28(8): 2404–2417. Publisher: SAGE Publications Ltd STMdoi: 10.1177/0962280218782293

16. Cox DR. Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological)* 1972; 34(2): 187–220. Publisher: [Royal Statistical Society, Wiley].

17. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. Wiley . 2002. Google-Books-ID: pivvAAAAMAAJ.

18. Laird NM, Ware JH. Random-Effects Models for Longitudinal Data. *Biometrics* 1982; 38(4): 963–974. Publisher: [Wiley, International Biometric Society]doi: 10.2307/2529876

19. Gordon PH, Cheng B, Salachas F, et al. Progression in ALS is not linear but is curvilinear. *Journal of Neurology* 2010; 257(10): 1713–1717. doi: 10.1007/s00415-010-5609-1

20. McCulloch CE, Searle SR, Neuhaus JM. *Generalized, Linear, and Mixed Models*. Wiley . 2008. Google-Books-ID: iDg0QwAACAAJ.

21. Schiratti JB, Allassonnière S, Routier A, Colliot O, Durrleman S. A Mixed-Effects Model with Time Reparametrization for Longitudinal Univariate Manifold-Valued Data. In: Ourselin S, Alexander DC, Westin CF, Cardoso MJ. , eds. *Information Processing in Medical Imaging*Lecture Notes in Computer Science. Springer International Publishing; 2015; Cham: 564–575

22. Schiratti JB. Methods and algorithms to learn spatio-temporal changes from longitudinal manifold-valued observations. 2017: 200.

23. Jedynak BM, Lang A, Liu B, et al. A Computational Neurodegenerative Disease Progression Score: Method and Results with the Alzheimer's Disease Neuroimaging Initiative Cohort. *NeuroImage* 2012; 63(3): 1478–1486. doi: 10.1016/j.neuroimage.2012.07.059

24. Bilgel M, Prince JL, Wong DF, Resnick SM, Jedynak BM. A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging. *NeuroImage* 2016; 134: 658–670. doi: 10.1016/j.neuroimage.2016.04.001

25. Bilgel M, Jedynak BM, Initiative ADN. Predicting time to dementia using a quantitative template of disease progression. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 2019; 11(1): 205–215. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1016/j.dadm.2019.01.005doi: 10.1016/j.dadm.2019.01.005

26. Raket LL. Statistical Disease Progression Modeling in Alzheimer Disease. *Frontiers in Big Data* 2020; 3.

27. Li D, Iddi S, Thompson WK, Donohue MC, for the Alzheimer's Disease Neuroimaging Initiative . Bayesian latent time joint mixed effect models for multicohort longitudinal data. *Statistical Methods in Medical Research* 2017; 28(3): 835–845. doi: 10.1177/0962280217737566

28. Iddi S, Li D, Aisen PS, et al. Estimating the evolution of disease in the Parkinson's Progression Markers Initiative. *Neuro-degenerative diseases* 2018; 18(4): 173–190. doi: 10.1159/000488780

29. Marinescu RV, Oxtoby NP, Young AL, et al. TADPOLE Challenge: Accurate Alzheimer's disease prediction through crowdsourced forecasting of future data. *PRedictive Intelligence in MEdicine. PRIME (Workshop)* 2019; 11843: 1–10. doi: 10.1007/978-3-030-32281-6_1

30. Koval I, Bône A, Louis M, et al. AD Course Map charts Alzheimer's disease progression. *Scientific Reports* 2021; 11(1): 8020. doi: 10.1038/s41598-021-87434-1

31. Proust-Lima C, Philipps V, Liquet B. Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm. *Journal of Statistical Software* 2017; 78: 1–56. doi: 10.18637/jss.v078.i02

32. Proust-Lima C, Saulnier T, Philipps V, et al. Describing complex disease progression using joint latent class models for multivariate longitudinal markers and clinical endpoints. *Statistics in Medicine* 2023; n/a(n/a). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.9844doi: 10.1002/sim.9844

33. Rizopoulos D. The *R* Package **JMbayes** for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC. *Journal of Statistical Software* 2016; 72(7). doi: 10.18637/jss.v072.i07

34. Leung KM, Elashoff RM, Afifi AA. CENSORING ISSUES IN SURVIVAL ANALYSIS. *Annual Review of Public Health* 1997; 18(1): 83–104. doi: 10.1146/annurev.publhealth.18.1.83

35. Koval I. Learning Multimodal Digital Models of Disease Progression from Longitudinal Data: Methods & Algorithms for the Description, Prediction and Simulation of Alzheimer's Disease Progression. 2020: 181.

36. Kuhn E, Lavielle M. Coupling a stochastic approximation version of EM with an MCMC procedure. *ESAIM: Probability and Statistics* 2004; 8: 115–131. Publisher: EDP Sciencesdoi: 10.1051/ps:2004007

37. Robbins H, Monro S. A Stochastic Approximation Method. *The Annals of Mathematical Statistics* 1951; 22(3): 400–407. Publisher: Institute of Mathematical Statisticsdoi: 10.1214/aoms/1177729586

38. Virtanen P, Gommers R, Oliphant TE, et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* 2020; 17: 261–272. doi: 10.1038/s41592-019-0686-2

39. Draper NR, Smith H. *Applied Regression Analysis*. John Wiley & Sons . 1998. Google-Books-ID: d6NsDwAAQBAJ.

40. Frost C, Thompson SG. Correcting for Regression Dilution Bias: Comparison of Methods for a Single Predictor Variable. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 2000; 163(2): 173–189. Publisher: [Wiley, Royal Statistical Society].

41. Spearman C. The Proof and Measurement of Association between Two Things. *The American Journal of Psychology* 1904; 15(1): 72–101. Publisher: University of Illinois Pressdoi: 10.2307/1412159

42. Fuller WA. *Measurement Error Models*. John Wiley & Sons . 2009. Google-Books-ID: Nalc0DkAJRYC.

43. Rosner B, Spiegelman D, Willett WC. Correction of Logistic Regression Relative Risk Estimates and Confidence Intervals for Random Within-Person Measurement Error. *American Journal of Epidemiology* 1992; 136(11): 1400–1413. doi: 10.1093/oxfordjournals.aje.a116453

44. Davidson-Pilon C. lifelines, survival analysis in Python.; 2023

45. Murawska M, Rizopoulos D, Lesaffre E. A Two-Stage Joint Model for Nonlinear Longitudinal Response and a Time-to-Event with Application in Transplantation Studies. *Journal of Probability and Statistics* 2012; 2012: e194194. Publisher: Hindawidoi: 10.1155/2012/194194

46. Miano B, Stoddard GJ, Davis S, Bromberg MB. Inter-evaluator reliability of the ALS functional rating scale. *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders: Official Publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases* 2004; 5(4): 235–239. doi: 10.1080/14660820410021302

47. Pan S, Hout v. dA. Bivariate joint models for survival and change of cognitive function. *Statistical Methods in Medical Research* 2023; 32(3): 474–492. Publisher: SAGE Publications Ltd STMdoi: 10.1177/09622802221146307

48. Kueffner R, Zach N, Bronfeld M, et al. Stratification of amyotrophic lateral sclerosis patients: a crowdsourcing approach. *Scientific Reports* 2019; 9: 690. doi: 10.1038/s41598-018-36873-4

49. Blanche P, Kattan MW, Gerds TA. The c-index is not proper for the evaluation of t-year predicted risks. *Biostatistics* 2019; 20(2): 347–357. doi: 10.1093/biostatistics/kxy006

50. Pölsterl S. scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn. *Journal of Machine Learning Research* 2020; 21(212): 1-6.

51. Atassi N, Berry J, Shui A, et al. The PRO-ACT database: design, initial analyses, and predictive features. *Neurology* 2014; 83(19): 1719–1725. doi: 10.1212/WNL.0000000000000951

52. Rooney J, Burke T, Vajda A, Heverin M, Hardiman O. What does the ALSFRS-R really measure? A longitudinal and survival analysis of functional dimension subscores in amyotrophic lateral sclerosis. *Journal of Neurology, Neurosurgery, and Psychiatry* 2017; 88(5): 381–385. doi: 10.1136/jnnp-2016-314661

53. Fournier CN, James V, Glass JD. Clinically meaningful change: evaluation of the Rasch-built Overall Amyotrophic Lateral Sclerosis Disability Scale (ROADS) and the ALSFRS-R. *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* 2023; 24(3-4): 311–316. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/21678421.2022.2153607doi: 10.1080/21678421.2022.2153607

54. Steffen T, Seney M. Test-retest reliability and minimal detectable change on balance and ambulation tests, the 36-item short-form health survey, and the unified Parkinson disease rating scale in people with parkinsonism. *Physical Therapy* 2008; 88(6): 733–746. doi: 10.2522/ptj.20070214

55. Feeney J, Savva GM, O'Regan C, King-Kallimanis B, Cronin H, Kenny RA. Measurement Error, Reliability, and Minimum Detectable Change in the Mini-Mental State Examination, Montreal Cognitive Assessment, and Color Trails Test among Community Living Middle-Aged and Older Adults. *Journal of Alzheimer's disease: JAD* 2016; 53(3): 1107–1114. doi: 10.3233/JAD-160248

56. Ding C, Wu Y, Chen X, et al. Global, regional, and national burden and attributable risk factors of neurological disorders: The Global Burden of Disease study 1990–2019. *Frontiers in Public Health* 2022; 10.

57. Feigin VL, Nichols E, Alam T, et al. Global, regional, and national burden of neurological disorders, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology* 2019; 18(5): 459–480. Publisher: Elsevierdoi: 10.1016/S1474-4422(18)30499-X

58. Schiratti JB, Allassonniere S, Colliot O, Durrleman S. A Bayesian mixed-effects model to learn trajectories of changes from repeated manifold-valued observations. 2017.

59. Talbott EO, Malek AM, Lacomis D. The epidemiology of amyotrophic lateral sclerosis. *Handbook of Clinical Neurology* 2016; 138: 225–238. doi: 10.1016/B978-0-12-802973-2.00013-6

60. Chen D, Liu S, Kingsbury P, et al. Deep learning and alternative learning strategies for retrospective real-world clinical data. *npj Digital Medicine* 2019; 2(1): 1–5. Number: 1 Publisher: Nature Publishing Groupdoi: 10.1038/s41746-019-0122-0

61. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *Journal of Clinical Epidemiology* 2019; 110: 12–22. doi: 10.1016/j.jclinepi.2019.02.004

62. Valliani AAA, Ranti D, Oermann EK. Deep Learning and Neurology: A Systematic Review. *Neurology and Therapy* 2019; 8(2): 351–365. doi: 10.1007/s40120-019-00153-8

63. Deuschl G, Beghi E, Fazekas F, et al. The burden of neurological diseases in Europe: an analysis for the Global Burden of Disease Study 2017. *The Lancet Public Health* 2020; 5(10): e551–e567. Publisher: Elsevierdoi: 10.1016/S2468-2667(20)30190-0

64. Prentice RL. Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model. *Biometrika* 1982; 69(2): 331–342. Publisher: [Oxford University Press, Biometrika Trust]doi: 10.2307/2335407

65. Gelman A, Rubin DB. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* 1992; 7(4): 457–472. Publisher: Institute of Mathematical Statistics.

66. Delyon B, Lavielle M, Moulines E. Convergence of a Stochastic Approximation Version of the EM Algorithm. *The Annals of Statistics* 1999; 27(1): 94–128. Publisher: Institute of Mathematical Statistics.

67. Allassonnière S, Kuhn E. Stochastic algorithm for Bayesian mixture effect template estimation. *ESAIM: Probability and Statistics* 2010; 14: 382–408. Publisher: EDP Sciencesdoi: 10.1051/ps/2009001

68. Allassonnière S, Kuhn E. Convergent stochastic Expectation Maximization algorithm with efficient sampling in high dimension. Application to deformable template model estimation. *Computational Statistics & Data Analysis* 2015; 91: 4–19. doi: 10.1016/j.csda.2015.04.011

69. Robbins H, Siegmund D. A CONVERGENCE THEOREM FOR NON NEGATIVE ALMOST SUPERMARTINGALES AND SOME APPLICATIONS**Research supported by NIH Grant 5-R01-GM-16895-03 and ONR Grant N00014-67-A-0108-0018.. In: Rustagi JS., ed. *Optimizing Methods in Statistics*Academic Press. 1971 (pp. 233–257)

70. Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data Via the EM Algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)* 1977; 39(1): 1–22. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.2517-6161.1977.tb01600.xdoi: 10.1111/j.2517-6161.1977.tb01600.x

71. Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models. *Statistics in Medicine* 2005; 24(11): 1713–1723. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.2059doi: 10.1002/sim.2059

72. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* 2016; 15(2): 155. Publisher: Elsevierdoi: 10.1016/j.jcm.2016.02.012

73. Murtaugh PA, Dickson ER, Van Dam GM, et al. Primary biliary cirrhosis: Prediction of short-term survival based on repeated patient visits. *Hepatology* 1994; 20(1): 126–134. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hep.1840200120doi: 10.1002/hep.1840200120

74. Lavielle M, Aarons L. What do we mean by identifiability in mixed effects models?. *Journal of Pharmacokinetics and Pharmacodynamics* 2016; 43(1): 111–122. doi: 10.1007/s10928-015-9459-4

75. Nunez OG, Concordet D. When is a nonlinear mixed-effects model identifiable ?.

76. Hougaard P. Frailty models for survival data. *Lifetime Data Analysis* 1995; 1(3): 255–273. doi: 10.1007/BF00985760

77. Andersen PK, Perme MP. Pseudo-observations in survival analysis. *Statistical Methods in Medical Research* 2010; 19(1): 71–99. doi: 10.1177/0962280209105020

Table 1 Absolute clinically meaningful estimated parameters bias of the Proposed model and reference models on real-like dataset

Legend: Proposed: the proposed joint model with latent disease age, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, Longitudinal: existing longitudinal model with latent disease age, AFT: Accelerated Failure Time model, JMbayes2: joint model with shared random effects.
Results are presented with the mean (SD) over the 10-fold cross-validation. P-values are computed using a Wilcoxon signed-rank test with Bonferroni correction between the Proposed model and each reference models. Results in bold are the best for the clinically meaningful estimated parameter. Values of the clinically meaningful estimated parameters that enabled the simulation, from which the biases were computed: -12.97 (0.11) points of ALSFRSr per year for the growth, 2.97 (0.02) years for the midpoint, 4.57 (0.03) years for the median.

	Proposed	Two stages	p-value	Longitudinal	p-value	AFT	p-value	JMbayes2	p-value
growth (pt/y)	0.59 (0.08)	-	-	**0.39 (0.08)**	3.9e-03	-	-	1.71 (0.12)	3.9e-03
midpoint (y)	0.06 (0.01)	-	-	**0.04 (0.01)**	3.9e-03	-	-	0.43 (0.04)	3.9e-03
median (y)	**0.09 (0.04)**	0.14 (0.05)	3.2e-01	-	-	0.80 (0.08)	5.9e-03	2.34 (0.14)	5.9e-03

Table 2 Prediction metrics of the Proposed model and reference models on real-like dataset
Legend: Proposed: the proposed joint model with latent disease age, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, Longitudinal: existing longitudinal model, AFT: Accelerated Failure Time model, JMbayes2: joint model with shared random effects.
Results are presented with mean (SD) over the 10-fold cross-validation. P-values are computed using a Wilcoxon signed-rank test or pairwise t-test with Bonferroni between the Proposed model and each reference model. IBS stand for Integrated Brier Score. ↓means that the metric should be minimised and ↑maximised. Results in bold are the best for each metric. 6,382 longitudinal predictions were made at 0.58 (0.41) years from the last visit.

	Proposed	Two stages	p-value	Longitudinal	p-value	AFT	p-value	JMbayes2	p-value
Absolute bias ↓	3.74 (3.80)	-	-	**3.73 (3.82)**	3.3e-05	-	-	4.09 (4.10)	3.9e-27
IBS ↓	**0.08 (0.02)**	0.08 (0.02)	3.2e-01	-	-	0.11 (0.03)	5.3e-05	0.11 (0.03)	7.5e-05
Mean AUC (1y, 1.5y) ↑	0.74 (0.10)	**0.75 (0.09)**	1.0e+00	-	-	0.60 (0.08)	2.2e-03	0.60 (0.09)	3.3e-03
C-index 1.0y ↑	**0.73 (0.07)**	0.73 (0.08)	1.0e+00	-	-	0.60 (0.07)	3.1e-03	0.59 (0.07)	2.5e-03
C-index 1.5y ↑	**0.73 (0.07)**	0.73 (0.07)	1.0e+00	-	-	0.60 (0.07)	2.0e-03	0.59 (0.07)	1.9e-03

Table 3 Characteristics of the PRO-ACT data
Legend: Results are presented with mean (SD) [class%]. There were no missing values in the dataset due to patient selection.

Characteristics	Values
Number of patients	2,528
Number of visits	23,143
Number of patients years	2,531
Percentage of censored events (%)	76.74
Number of visits per patients	9.2 (4.3)
Time of follow-up (years)	1.0 (0.6)
Time between visits (months)	1.5 (0.9)
Gender (Male)	1,575 [62.3 %]
Symptom onset (Spinal)	1,952 [77.2 %]
Age at first symptoms	54.0 (11.3)
Time from first symptoms to baseline (years)	1.6 (0.9)
ALSFRSr total (baseline)	37.9 (5.4)

Table 4 Clinically meaningful estimated parameters bias of the Proposed model and reference models on real-like dataset

Legend: Proposed: the proposed joint model with latent disease age, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, Longitudinal: existing longitudinal model with latent disease age, AFT: Accelerated Failure Time model, JMbayes2: joint model with shared random effects.
Results are presented with the mean (SD) over the 10-fold cross-validation. P-values are computed using a Wilcoxon signed-rank test with Bonferroni correction between the Proposed model and each of the reference models.

	Proposed	Two stages	p-value	Longitudinal	p-value	AFT	p-value	JMbayes2	p-value
growth (pt/y)	-14.17 (0.08)	-	-	-13.95 (0.08)	9.8e-11	-	-	-15.35 (0.10)	4.6e-14
midpoint (y)	2.78 (0.01)	-	-	2.79 (0.01)	1.6e-04	-	-	3.22 (0.01)	1.8e-13
median (y)	4.15 (0.03)	4.14 (0.02)	6.2e-01	-	-	4.69 (0.02)	4.4e-11	4.47 (0.09)	2.0e-06

Table 5 Prediction metrics of the Proposed model and reference models on PRO-ACT data
Legend: Proposed: the proposed joint model with latent disease age, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, Longitudinal: existing longitudinal model with latent disease age, AFT: Accelerated Failure Time model, JMbayes2: joint model with shared random effects.
Results are presented with the mean (SD) over the 10-fold cross-validation. P-values are computed using a Wilcoxon signed-rank test or pairwise t-test with Bonferroni between the Proposed model and each of the reference models. IBS stand for Integrated Brier Score. ↓means that the metric should be minimised and ↑maximised. Results in bold are the best for each metric. 18,077 longitudinal predictions were made at 0.63 (0.55) years from the last visit.

	Proposed	Two stages	p-value	Longitudinal	p-value	AFT	p-value	JMbayes2	p-value
Absolute bias ↓	4.21 (4.41)	-	-	**4.18 (4.38)**	2.2e-73	-	-	4.24 (4.14)	1.4e-17
IBS ↓	**0.10 (0.01)**	0.11 (0.01)	2.5e-04	-	-	0.12 (0.01)	1.1e-04	0.10 (0.01)	1.0e+00
Mean AUC (1y, 1.5y)	**0.67 (0.07)**	0.62 (0.08)	2.8e-03	-	-	0.42 (0.07)	4.3e-05	0.61 (0.09)	1.7e-03
C-index 1.0y ↑	**0.69 (0.05)**	0.63 (0.06)	4.0e-04	-	-	0.41 (0.05)	1.1e-06	0.63 (0.06)	3.1e-04
C-index 1.5y ↑	**0.70 (0.05)**	0.65 (0.05)	9.7e-04	-	-	0.41 (0.05)	1.6e-06	0.66 (0.05)	1.8e-03

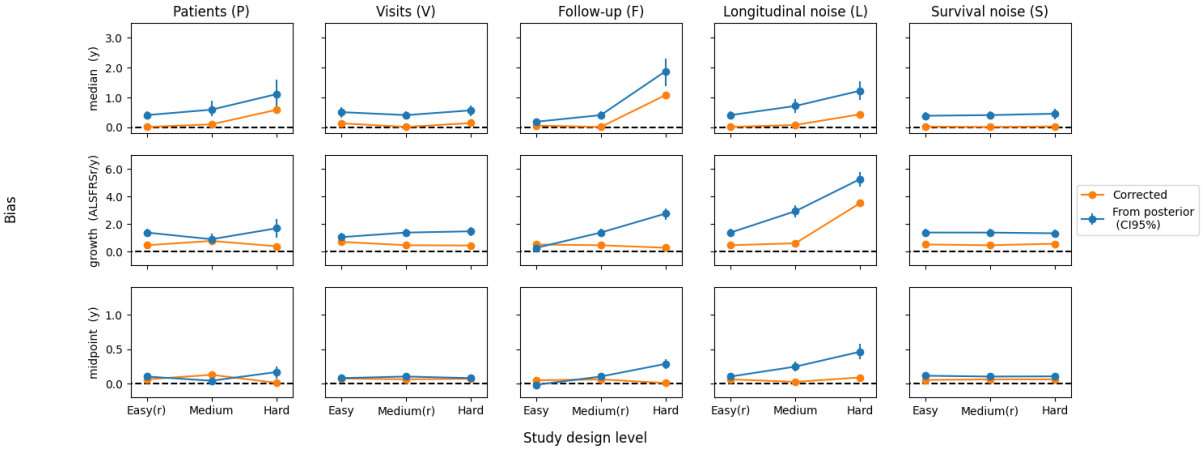


Figure 1 Absolute estimated bias for clinically meaningful estimated parameters (median, growth and midpoint) on all simulated datasets

Legend: Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. Easy, Medium and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in part 3.1.2 and summarised in the table F3 in annexe. Datasets with (r) correspond to the real-like dataset.

corrected: mean bias extracted from Robbin Monroe algorithm corrected for attenuation bias, from posterior: mean bias with CI 95% of the posterior MCMC chain. Values of the clinically meaningful estimated parameters that enabled the simulation, from which the bias is computed, were 4.48 (0.12) years for the median, -13.19 (0.21) point of ALSFRSr per year for the growth, 2.92 (0.06) years for the midpoint.

□

APPENDIX

A LIKELIHOOD TOTAL

$$
\begin{aligned}
& \log p((y, T_e, B_e), z, \theta \mid \Pi) \\
= & \sum_{i,j} -\log\left(\sigma\sqrt{2\pi}\right) - \frac{1}{2\sigma^2}\left(y_{i,j} - \gamma_0\left(\psi_i\left(t_{i,j}\right)\right)\right)^2 \\
+ & \sum_i \mathbb{1}_{B_{e_i}} \times \log\left(h_i(t_{e_i})\right) + \sum_i \log\left(S_i(t_{e_i})\right) \\
- & \log\left(\sigma_{\tilde{g}}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{g}}^2}\left(\tilde{g} - \overline{\tilde{g}}\right)^2 \\
- & \log\left(\sigma_{\tilde{v}_0}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{v}_0}^2}\left(\tilde{v}_0 - \overline{\tilde{v}_0}\right)^2 \\
- & \log\left(\sigma_{\tilde{v}}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{v}}^2}\left(\tilde{v} - \overline{\tilde{v}}\right)^2 \\
- & \log\left(\sigma_{\tilde{\rho}}\sqrt{2\pi}\right) - \frac{1}{2\sigma_{\tilde{\rho}}^2}\left(\tilde{\rho} - \overline{\tilde{\rho}}\right)^2 \\
- & N\log\left(\sigma_\tau\sqrt{2\pi}\right) - \frac{1}{2\sigma_\tau^2}\sum_i(\tau_i - t_0)^2 \\
- & N\log\left(\sigma_\xi\sqrt{2\pi}\right) - \frac{1}{2\sigma_\xi^2}\sum_i(\xi_i - \overline{\xi})^2
\end{aligned}
$$

B SUFFICIENT STATISTICS

The convergence of the Monte-Carlo Markov Chain Stochastic Approximation Expectation-Maximization (MCMC-SAEM) algorithm has been proven in [36] for models which lie into the curved exponential family. For such a family of distributions, the log-likelihood can be written as:

$$\log\ p(\mathbf{Y}, \mathbf{z}, \theta, \Pi) = -\Phi(\theta, \Pi) + \langle S(\mathbf{Y}, \mathbf{z}), f(\theta, \Pi)\rangle$$

where Φ and f are smooth functions, and S are called the sufficient statistics. The sufficient statistics are to be understood as a summary of the required information from the latent variables $\mathbf{z}$ and the observations $\mathbf{Y}$ in our case (y, t_e, B). Our model falls in such a category and sufficient statistics are described below.

The idea is to rewrite likelihood in the above form to get sufficient statistics. As a reminder, note that there are N patients indexed by i and each has n_i visits indexed by j.

$$
\begin{aligned}
\log q((y, T_e, B_e), z, \theta \mid \Pi) = \sum_{i,j} -\log\left(\sigma\sqrt{2\pi}\right) &- \underbrace{\langle [\|y_{ij}\|^2]_{ij}}_{S_1(Y,z)} \underbrace{-2[y_{ij}^T \gamma_0\left(\psi_i\left(t_{i,j}\right)\right)]_{ij}}_{S_2(Y,z)} + \underbrace{[\|\gamma_0\left(\psi_i\left(t_{i,j}\right)\right)\|^2]_{ij}}_{S_3(Y,z)}, \frac{1}{2\sigma^2}\mathbf{1}_{\sum n_i}\rangle \\
&+ \underbrace{\langle [\mathbb{1}_{B_{e_i}} \times \log\left(h_i(t_{e_i})\right) + \sum_i \log\left(S_i(t_{e_i})\right)]_i, \mathbf{1}_N\rangle}_{S_4(Y,z)} \\
&- \ln(\sigma_{\tilde{g}}\sqrt{2\pi}) + \langle \underbrace{[\tilde{g}^2]}_{S_5(Y,z)}, -\frac{1}{2\sigma_{\tilde{g}}^2}\mathbf{1}_1\rangle + \langle \underbrace{[\tilde{g}]}_{S_6(Y,z)}, \frac{1}{\sigma_{\tilde{g}}^2}[\overline{\tilde{g}}]\rangle - \sum_{k=1}^{1} \frac{1}{2\sigma_{\tilde{g}}^2}\overline{\tilde{g}}^2 \\
&- \ln(\sigma_{\tilde{v}_0}\sqrt{2\pi}) + \langle \underbrace{[\tilde{v}_0^2]}_{S_7(Y,z)}, -\frac{1}{2\sigma_{\tilde{v}_0}^2}\mathbf{1}_1\rangle + \langle \underbrace{[\tilde{v}_0]}_{S_8(Y,z)}, \frac{1}{\sigma_{\tilde{v}_0}^2}[\overline{\tilde{v}_0}]\rangle - \sum_{k=1}^{1} \frac{1}{2\sigma_{\tilde{v}_0}^2}\overline{\tilde{v}_0}^2 \\
&- \ln(\sigma_{\tilde{v}}\sqrt{2\pi}) + \langle \underbrace{[\tilde{v}^2]}_{S_9(Y,z)}, -\frac{1}{2\sigma_{\tilde{v}}^2}\mathbf{1}_1\rangle + \langle \underbrace{[\tilde{v}]}_{S_{10}(Y,z)}, \frac{1}{\sigma_{\tilde{v}}^2}[\overline{\tilde{v}}]\rangle - \sum_{k=1}^{1} \frac{1}{2\sigma_{\tilde{v}}^2}\overline{\tilde{v}}^2 \\
&- \ln(\sigma_{\tilde{\rho}}\sqrt{2\pi}) + \langle \underbrace{[\tilde{\rho}^2]}_{S_{11}(Y,z)}, -\frac{1}{2\sigma_{\tilde{\rho}}^2}\mathbf{1}_1\rangle + \langle \underbrace{[\tilde{\rho}]}_{S_{12}(Y,z)}, \frac{1}{\sigma_{\tilde{\rho}}^2}[\overline{\tilde{\rho}}]\rangle - \sum_{k=1}^{1} \frac{1}{2\sigma_{\tilde{\rho}}^2}\overline{\tilde{\rho}}^2 \\
&- N\log(\sigma_\tau\sqrt{2\pi}) + \langle \underbrace{[\tau_i^2]_i}_{S_{13}(Y,z)}, -\frac{1}{2\sigma_\tau^2}\mathbf{1}_N\rangle + \langle \underbrace{[\tau_i]_i}_{S_{14}(Y,z)}, \frac{1}{\sigma_\tau^2}\overline{\tau}\mathbf{1}_N\rangle - \frac{1}{2\sigma_\tau^2}N\overline{\tau}^2 \\
&- N\log(\sigma_\xi\sqrt{2\pi}) + \langle \underbrace{[\xi_i^2]_i}_{S_{15}(Y,z)}, -\frac{1}{2\sigma_\xi^2}\mathbf{1}_N\rangle + \langle \underbrace{[\xi_i]_i}_{S_{16}(Y,z)}, \frac{1}{\sigma_\xi^2}\overline{\xi}\mathbf{1}_N\rangle - \frac{1}{2\sigma_\xi^2}N\overline{\xi}^2
\end{aligned}
$$

C MAXIMIZATION UPDATE RULES

To find the update rule of the different parameters, we need to find the new parameter θ that maximizes the log-likelihood. As expressions are convex in θ we can simply derive and look for a critical point. We derive the log-likelihood with respect to each maximised fixed effect. Note that only maximised fixed effects are updated by a maximization rule, other parameters are latent variables that are sampled. $\overline{\xi}$ is first maximised and then set to 0. As a reminder, note that there are N patients indexed by i and that each of them has n_i visits indexed by j. At iteration k, we can use $\tilde{S}^{(k+1)}$ computed with the parameters at iteration k and the formula of $S(Y, z)$ to compute the parameters at iteration $(k + 1)$.

$$
\begin{aligned}
(\sigma^2)^{(k+1)} &\leftarrow \frac{1}{N}[\tilde{S}_1^{(k)} - 2\tilde{S}_2^{(k)} + \tilde{S}_3^{(k)}]^T \mathbf{1}_1 \\
(\overline{\tilde{g}}_j)^{(k+1)} &\leftarrow \tilde{S}_6^{(k)} \\
(\overline{\tilde{\upsilon}_{0j}})^{(k+1)} &\leftarrow \tilde{S}_8^{(k)} \\
(\overline{\tilde{\nu}}_j)^{(k+1)} &\leftarrow \tilde{S}_{10}^{(k)} \\
(\overline{\tilde{\rho}}_j)^{(k+1)} &\leftarrow \tilde{S}_{12}^{(k)} \\
(\overline{\tau})^{(k+1)} &\leftarrow \frac{1}{N}\tilde{S}_{14}^{(k)} \\
(\sigma_\tau^2)^{(k+1)} &\leftarrow \frac{1}{p}[\tilde{S}_{13}^{(k)} - 2\overline{\tau}\tilde{S}_{14}^{(k)}]^T \mathbf{1}_N + \overline{\tau}^2 \\
(\overline{\xi})^{(k+1)} &\leftarrow \frac{1}{N}\tilde{S}_{16}^{(k)} \\
(\sigma_\xi^2)^{(k+1)} &\leftarrow \frac{1}{N}[\tilde{S}_{15}^{(k)} - 2\overline{\xi}\tilde{S}_{16}^{(k)}]^T \mathbf{1}_N + \overline{\xi}^2
\end{aligned}
$$

D REFERENCE MODELS EQUATIONS

Model	Inputs	Effects		Random effects structure	Link functions	
		Fix	Random	$\psi_i(t)$	Longitudinal $\gamma(t)$	Survival S(t)
Longitudinal	t	g, υ_0, t_0	ξ, τ	$e^{\xi}(t-\tau)+t_0$	$\left(1+g\times\exp(-\upsilon_0\frac{(g+1)^2}{g}(\psi_i(t)-t_0))\right)^{-1}$	-
AFT	t	ν_0, ρ_0	-	-	-	$exp\left(-\left(\frac{t}{\exp(\nu_0)}\right)^{\exp(\rho_0)}\right)$
Two-stages	$t, x=(\xi,\tau)$	ν_0, ν_1, ρ_0	-	-	-	$exp\left(-\left(\frac{t}{\exp(\nu_0+\nu_1 x)}\right)^{\exp(\rho_0)}\right)$
Proposed	t	$g, \upsilon_0, t_0, \rho, \nu$	ξ, τ	$e^{\xi}(t-\tau)+t_0$	$\left(1+g\times\exp(-\upsilon_0\frac{(g+1)^2}{g}(\psi_i(t)-t_0))\right)^{-1}$	$exp\left(-\left(\frac{\psi_i(t)}{\nu}\right)^{\rho}\right)$
JMBayes2	t	$\beta_0, \beta_1, \alpha, spl(t)$	$b_{i,0}, b_{i,1}$	$(\beta_0+b_{i,0})+(\beta_1+b_{i,1})t$	$\left(1+g\times\exp(\psi_i(t))\right)^{-1}$	$\exp\left(-\int_0^t \exp(spl(u)+\alpha\gamma(u))du\right)$

Table D1 Specification of reference models

Legend: Longitudinal: existing longitudinal model with latent disease age, Two-stage model: AFT survival model that used as covariate random effects of the Longitudinal model, AFT: Accelerated Failure Time model, Proposed: the proposed joint model with latent disease age, JMbayes2: joint model with shared random effects. $spl(t)$: spline function, $x=(\xi,\tau)$ are extracted from the Longitudinal model and used as covariates,

Model	Survival	Longitudinal	
	Median	Growth	Midpoint
Longitudinal	-	$v_0\frac{(g+1)^2}{4g}$	$\log(g)\frac{g}{v_0(g+1)^2}+t_0$
Two-stages	available in package	-	-
AFT	available in package	-	-
Proposed	$\nu\ln(2)^{\frac{1}{\rho}}$	$v_0\frac{(g+1)^2}{4g}$	$\log(g)\frac{g}{v_0(g+1)^2}+t_0$
JMBayes2	from S(t) estimation	$-\beta_1/4$	$-\beta_0$

Table D2 Clinically meaningful estimated parameters formulas

Legend: Longitudinal: existing longitudinal model with latent disease age, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, AFT: Accelerated Failure Time model, Proposed: the proposed joint model with latent disease age, JMbayes2: joint model with shared random effects.

If a function to extract the clinically meaningful estimated parameters was available in the package it was directly used and noticed as available in the package. For the JMBayes2 package, we estimated the function S(t) to find the median of survival (from S(t) estimation).

E ANALYSIS OF PRO-ACT FOR REAL-LIKE SIMULATION

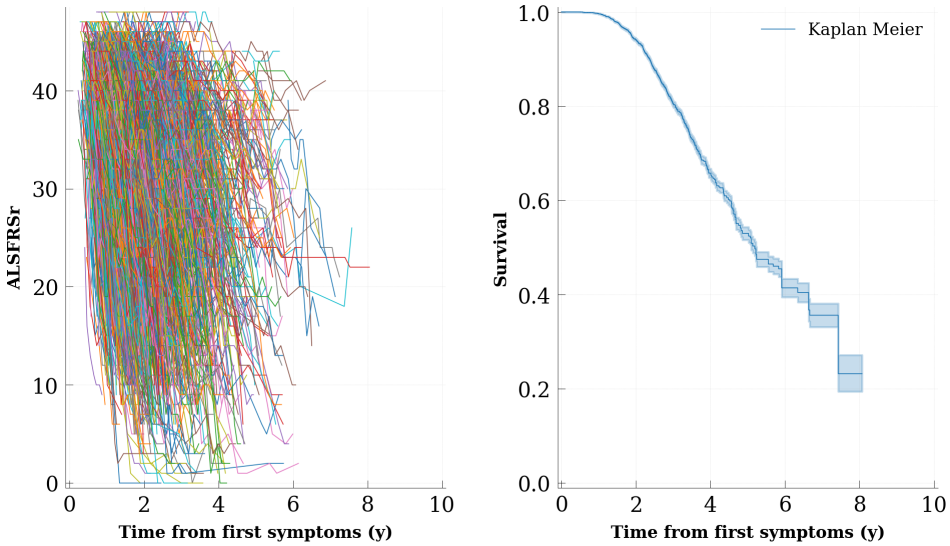


Figure E1 PRO-ACT longitudinal and survival data
Legend: The spaghetti plot represents each patient trajectory of ALSFRSr from first symptoms. Survival corresponds to time to death or tracheotomy.

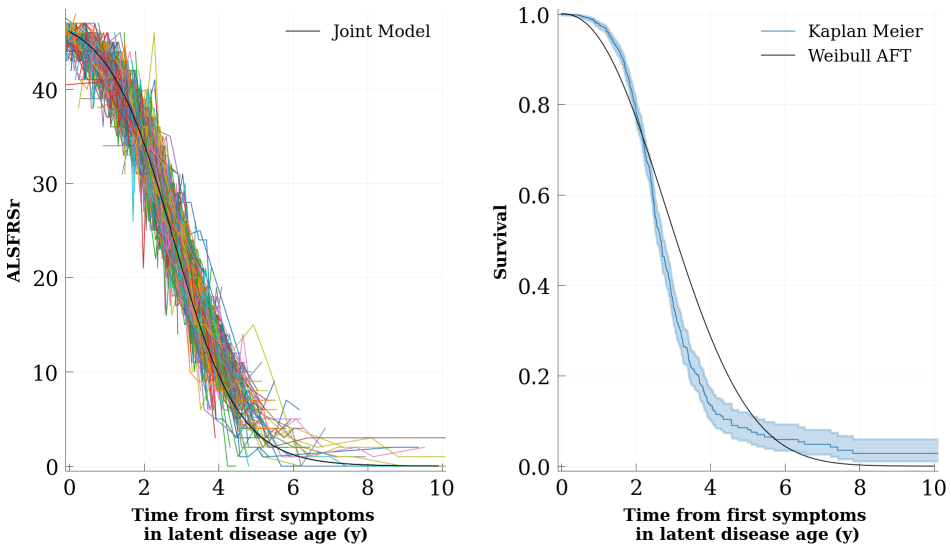


Figure E2 PRO-ACT longitudinal and survival data reparametrized in latent disease age
Legend: The spaghetti plot represents each patient trajectory of ALSFRSr from first symptoms in latent disease age. Survival corresponds to time to death or tracheotomy in latent disease age.

F SUMMARY OF PARAMETERS FOR SIMULATION

Table F3 Data simulation parameters by scenario

Legend: Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. (r) indicate when ALS real-like parameters are used. For each scenario, three datasets are simulated [Easy, Medium, Hard] that correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in part 3.1.2 and summarised in the table F3 in annexe. Note that the population estimated reference time, t_0, correspond also to the mean time shift.

Type	Parameters Name		ALS Symbol	Patient real-like (r)	Visit number (P)	Follow up density(V)	Longitudinal (F)	Survival noise (L)	noise (S)
Patients	Patient number		N	1,000	[r, 500, 200]	r	r	r	r
Random Effect	Time shift	(std)	σ_τ	1.04	r	r	r	r	r
	Individual log-rate factor	(mean)	$\overline{\xi}$	0	r	r	r	r	r
		(std)	σ_ξ	0.73	r	r	r	r	r
Longitudinal Fixed Effects	Population estimated reference time		t_0	1.17	r	r	r	r	r
	Speed of the logistic curve		v_0	1.13	r	r	r	r	r
	Curve value at t_0: $\frac{1}{1+g}$		g	6.40	r	r	r	r	r
	Estimated noise		σ	0.04	r	r	r	[r, 0.1, 0.2]	r
Survival Fixed Effects	Scale of the Weibull distribution		ν	3.62	r	r	r	r	[3.11, r, 3.84]
	Shape of the Weibull distribution		ρ	2.25	r	r	r	r	[3, r, 2]
Visits	Time between τ and baseline	(mean)	$\overline{\delta_f}$	0.4	r	r	r	r	r
		(std)	σ_{δ_f}	0.84	r	r	r	r	r
	Time of follow up	(mean)	$\overline{T_f}$	0.96	r	r	[2, r, 0.5]	r	r
		(std)	σ_{T_f}	0.87	r	r	r	r	r
	Time between visits	(mean)	$\overline{\delta_v}$	1.47	r	[1, r, 3]	r	r	r
		(std)	σ_{δ_v}	0.5	r	r	r	r	r

G SIMULATED DATA STATISTICS

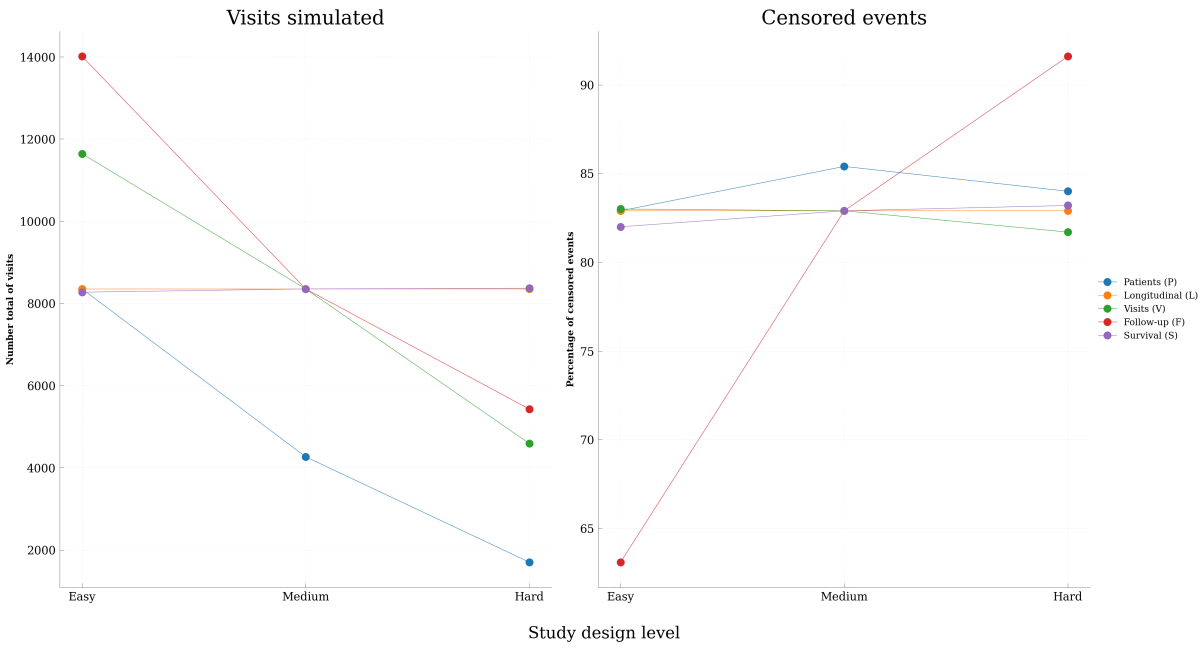


Figure G3 Statistics on simulated data depending on the scenario
Legend: Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. Easy, Medium and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in part 3.1.2 and summarised in the table F3 in annexe. Note that the Longitudinal (L) and the Survival (S) scenarios have almost the same number of visits. The Visits (V) and Longitudinal (L) scenarios have almost the same censoring rate for the easiest configuration, and the Survival (S) and Longitudinal (L) scenarios for the most difficult configuration.

H VALIDATION - POSTERIOR EFFECTS ESTIMATION

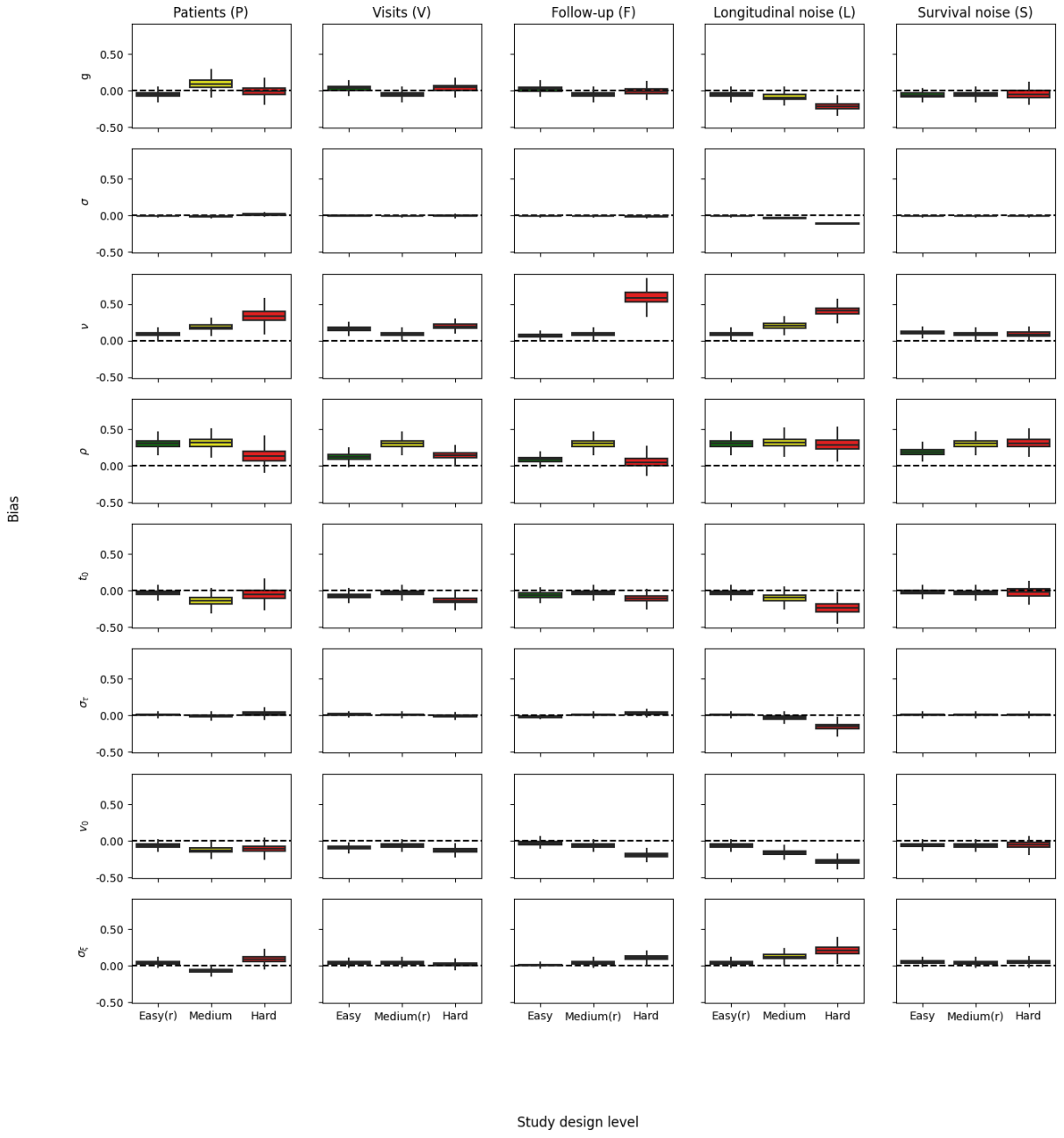


Figure H4 Fixed effects estimation bias from posterior distributionon simulated datasets

Legend: In ordinate the normalised bias on the true value of the parameters, in abscisse the different dataset for each scenario.

Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. With (r) that indicates when the study design corresponds to the real-like design. Easy, Medium and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in part 3.1.2 and summarised in the table F3 in annexe.

g: Curve value at t_0: $\frac{1}{1+g}$, σ: Estimated noise, ν: Scale of the Weibull distribution, ρ: Shape of the Weibull distribution, t_0: Population estimated reference time (mean time shift), σ_τ: Time-shift standard-deviation, v_0: Speed of the logistic curve, σ_ξ: Individual log-rate factor (the mean is fixed to 0 to ensure identifiability).

The distribution was computed using 30,000 MCMC iterations.

Table H4 Intra-class correlation of random effects from the mean of posterior distribution on simulated datasets

Legend: Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. Easy, Medium and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in part 3.1.2 and summarised in the table F3 in annexe.

τ: the individual time-shit, ξ: the progression log-rate factor. Intra-class correlations are computed between the random effects estimated and the one simulated.

		Patients (P)		Visits (V)		Follow-up (F)		Longitudinal noise (L)		Survival noise (S)	
		average	CI95%	average	CI95%	average	CI95%	average	CI95%	average	CI95%
ξ	Easy	0.91	[0.90 0.92]	0.93	[0.92 0.93]	0.97	[0.97 0.97]	0.91	[0.90 0.92]	0.91	[0.90 0.92]
	Medium	0.89	[0.87 0.91]	0.91	[0.90 0.92]	0.91	[0.90 0.92]	0.83	[0.81 0.85]	0.91	[0.90 0.92]
	Hard	0.93	[0.91 0.95]	0.90	[0.89 0.91]	0.86	[0.84 0.87]	0.75	[0.72 0.78]	0.91	[0.90 0.92]
τ	Easy	0.96	[0.95 0.96]	0.97	[0.96 0.97]	0.98	[0.97 0.98]	0.96	[0.95 0.96]	0.96	[0.95 0.96]
	Medium	0.94	[0.92 0.95]	0.96	[0.95 0.96]	0.96	[0.95 0.96]	0.88	[0.86 0.89]	0.96	[0.95 0.96]
	Hard	0.96	[0.95 0.97]	0.94	[0.93 0.94]	0.92	[0.91 0.93]	0.66	[0.62 0.70]	0.96	[0.95 0.96]

I VALIDATION - ATTENUATION BIAS CORRECTION

Table I5 Clinically meaningful estimated parameters signed bias on simulated data

Legend: Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. Easy, Medium and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in part 3.1.2 and summarised in the table F3 in annexe. bias: mean (SD) computed from the posterior distribution, corrected bias: the mean extracted from the Robbin Monroe algorithm corrected for attenuation bias, truth: the value used for data simulation.

		Patients (P)			Visits (V)			Follow-up (F)			Longitudinal noise (L)			Survival noise (S)		
		bias	corrected bias	truth	bias	corrected bias	truth	bias	corrected bias	truth	bias	corrected bias	truth	bias	corrected bias	truth
median	Easy	0.40 (0.07)	0.00	4.50	0.50 (0.08)	0.13	4.47	0.18 (0.05)	0.05	4.40	0.40 (0.07)	0.00	4.50	0.38 (0.06)	0.02	4.18
	Medium	0.59 (0.13)	0.09	4.76	0.40 (0.07)	0.00	4.50	0.40 (0.07)	0.00	4.50	0.71 (0.11)	0.07	4.50	0.40 (0.07)	0.00	4.50
	Hard	1.11 (0.21)	0.58	4.34	0.56 (0.09)	0.14	4.42	1.88 (0.24)	1.08	4.49	1.22 (0.15)	0.43	4.50	0.45 (0.08)	0.02	4.61
growth	Easy	1.37 (0.14)	0.45	-13.19	1.04 (0.13)	0.69	-13.22	0.25 (0.08)	0.50	-13.25	1.37 (0.14)	0.45	-13.19	1.37 (0.14)	0.50	-13.09
	Medium	0.89 (0.19)	0.76	-12.47	1.37 (0.14)	0.45	-13.19	1.37 (0.14)	0.45	-13.19	2.92 (0.21)	-0.59	-13.19	1.37 (0.14)	0.45	-13.19
	Hard	1.68 (0.32)	0.37	-13.49	1.46 (0.15)	0.43	-13.39	2.75 (0.19)	-0.26	-13.36	5.26 (0.24)	-3.51	-13.19	1.32 (0.14)	0.56	-13.22
midpoint	Easy	0.10 (0.02)	-0.06	2.93	0.08 (0.02)	-0.07	2.90	-0.02 (0.01)	-0.05	2.83	0.10 (0.02)	-0.06	2.93	0.11 (0.02)	-0.05	2.94
	Medium	0.04 (0.03)	-0.13	3.09	0.10 (0.02)	-0.06	2.93	0.10 (0.02)	-0.06	2.93	0.25 (0.03)	-0.02	2.93	0.10 (0.02)	-0.06	2.93
	Hard	0.17 (0.04)	0.01	2.80	0.08 (0.02)	-0.07	2.87	0.29 (0.03)	-0.01	2.93	0.46 (0.05)	0.09	2.93	0.10 (0.02)	-0.06	2.91

Table I6 Intra-class correlation of random effects from personalization on simulated datasets

Legend: Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. Easy, Medium and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in part 3.1.2 and summarised in the table F3 in annexe.

τ: the individual time-shit, ξ: the progression log-rate factor. Intra-class correlations are computed between the random effects estimated and the one simulated.

		Patients (P)		Visits (V)		Follow-up (F)		Longitudinal noise (L)		Survival noise (S)	
		average	CI95%	average	CI95%	average	CI95%	average	CI95%	average	CI95%
ξ	Easy	0.87	[0.86 0.89]	0.89	[0.87 0.90]	0.96	[0.95 0.96]	0.87	[0.86 0.89]	0.87	[0.85 0.88]
	Medium	0.83	[0.80 0.86]	0.87	[0.86 0.89]	0.87	[0.86 0.89]	0.75	[0.71 0.78]	0.87	[0.86 0.89]
	Hard	0.90	[0.87 0.93]	0.86	[0.85 0.88]	0.79	[0.76 0.81]	0.67	[0.63 0.71]	0.87	[0.86 0.89]
τ	Easy	0.97	[0.96 0.97]	0.97	[0.97 0.98]	0.98	[0.98 0.98]	0.97	[0.96 0.97]	0.97	[0.96 0.97]
	Medium	0.95	[0.94 0.96]	0.97	[0.96 0.97]	0.97	[0.96 0.97]	0.91	[0.90 0.92]	0.97	[0.96 0.97]
	Hard	0.96	[0.95 0.97]	0.96	[0.95 0.96]	0.95	[0.94 0.96]	0.71	[0.67 0.75]	0.97	[0.96 0.97]

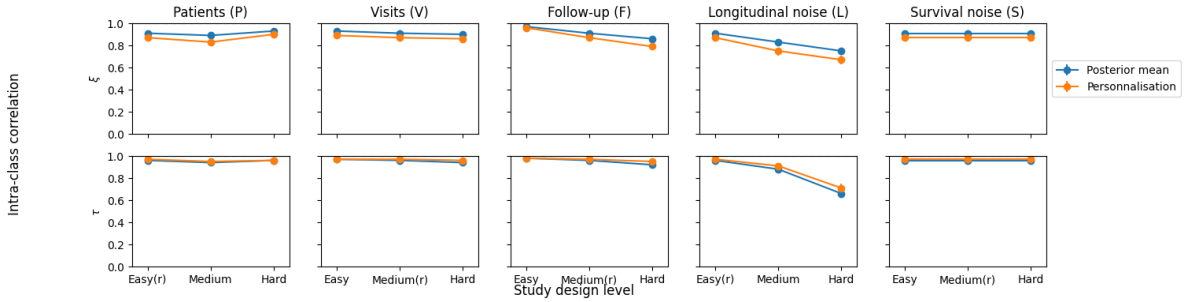


Figure I5 Intra-class correlation of random effects on simulated datasets

Legend: Results are presented with mean with CI95%.

In ordinate, the two random effects: τ the individual time-shit and ξ the progression log-rate factor. Intra-class correlations are computed between the random effects estimated and the one simulated.

In abscise the different dataset for each scenario: Patients(P): scenario with different number of patients, Visits (V): scenario with different density of visits, Follow-up (F): scenario with different follow-up duration, Longitudinal noise (L): scenario with different noises on longitudinal outcome, Survival noise (S): Scenario with different standard deviation on survival. Easy, Medium and Hard correspond to an increased difficulty of simulated data for the estimation, more information on simulation is available in part 3.1.2 and summarised in the table F3 in annexe. Datasets with (r) correspond to the real-like dataset.

J BENCHMARK - SIMULATED AND REAL DATA

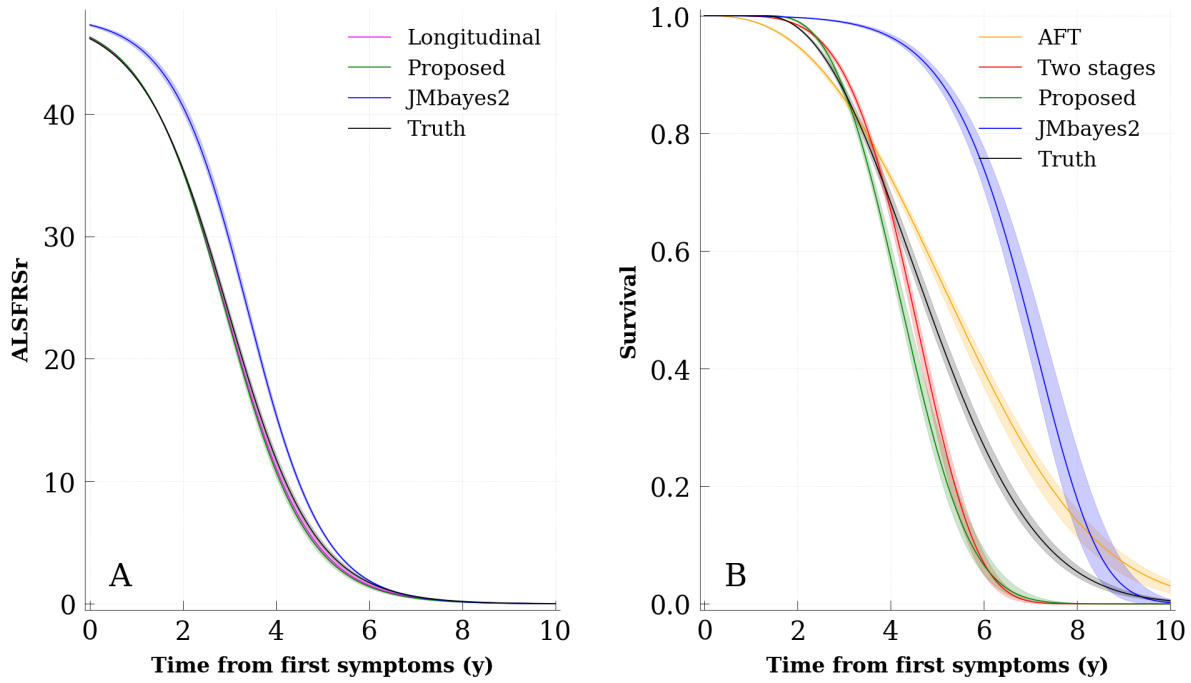


Figure J6 Average patient curve of the Proposed model and reference models on real-like dataset
Legend: Results are presented with mean over the 10-fold cross-validation and the maximum and minimum variation for each model.
Panel A: *Longitudinal: existing longitudinal model with latent disease age, Proposed: the proposed joint model with latent disease age, JMbayes2: joint model with shared random effects, Truth: real average patient that enabled data simulation,*
Panel B: *AFT: Accelerated Failure Time model, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, Proposed: the Proposed model, JMbayes2: joint model with shared random effects, Truth: real average patient that enabled data simulation.*
Note that for the longitudinal process, the curves of the Truth, the Longitudinal and Proposed models are superimposed.

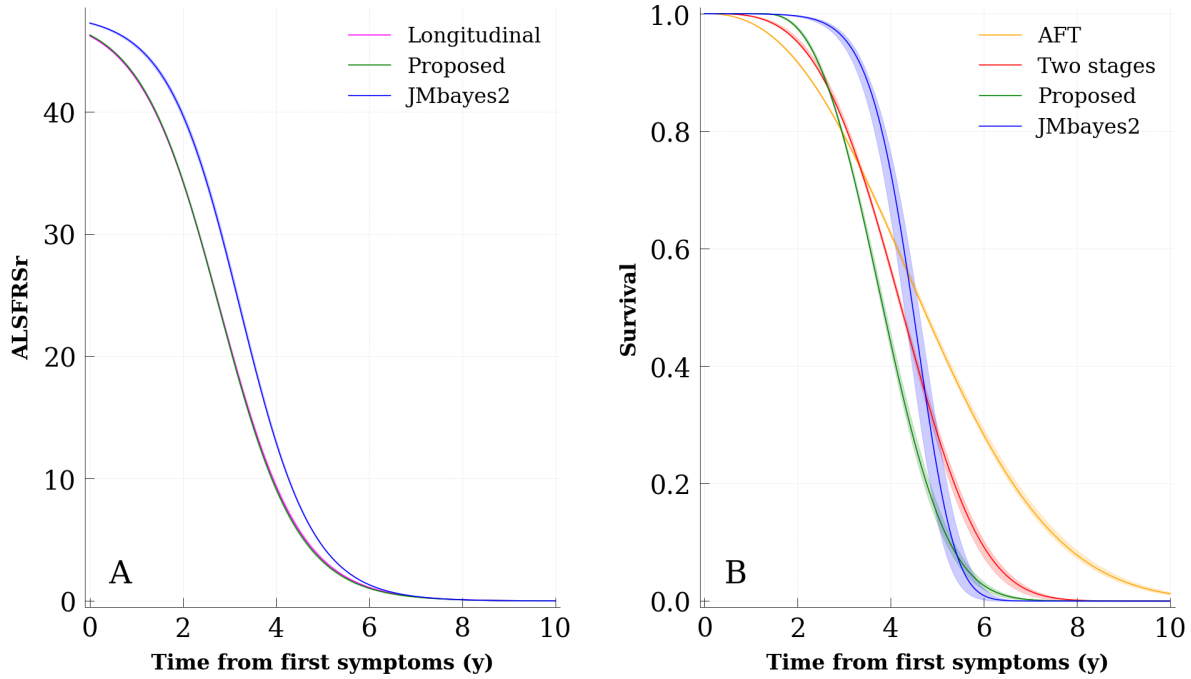


Figure J7 Average patient curve of the Proposed model and reference models on PRO-ACT dataset
Legend: Results are presented with mean over the 10-fold cross-validation and the maximum and minimum variation for each model.
Panel A: *Longitudinal: existing longitudinal model with latent disease age, Proposed: the proposed joint model with latent disease age, JMbayes2: joint model with shared random effects,*
Panel B: *AFT: Accelerated Failure Time model, Two-stage model: AFT survival model that uses as covariate random effects of the Longitudinal model, Proposed: the Proposed model, JMbayes2: joint model with shared random effects.*
Note that for the longitudinal process, the curves of the Longitudinal and Proposed models are superimposed.

Bibliography

Aalen, O. O. (Aug. 1989). "A linear regression model for the analysis of life times". eng. In: *Statistics in Medicine* 8.8, pp. 907–925. DOI: 10.1002/sim.4780080803.

Abi Nader, C. et al. (Jan. 2020). "Monotonic Gaussian Process for spatio-temporal disease progression modeling in brain imaging data". eng. In: *NeuroImage* 205, p. 116266. DOI: 10.1016/j.neuroimage.2019.116266.

Ad, d. J., v. E. Rpa, and v. d. B. Lh (Oct. 2019). "Evidence for a multimodal effect of riluzole in patients with ALS?" en. In: *Journal of neurology, neurosurgery, and psychiatry* 90.10. DOI: 10.1136/jnnp-2018-320211.

Aging, N. I. on (2023). *NIA-Funded Active Alzheimer's and Related Dementias Clinical Trials and Studies*. en.

Ai, U., H. Street, and S. Francisco (2021). "PROCOVA™ Handbook for the Target Trial Statistician". en. In.

Allison, P. (Mar. 2018). *For Causal Analysis of Competing Risks, Don't Use Fine & Gray's Subdistribution Method*. en-US.

Alsefri, M. et al. (Apr. 2020). "Bayesian joint modelling of longitudinal and time to event data: a methodological review". en. In: *BMC Medical Research Methodology* 20.1, p. 94. DOI: 10.1186/s12874-020-00976-2.

Andersen, P. K. and M. P. Perme (Feb. 2010). "Pseudo-observations in survival analysis". eng. In: *Statistical Methods in Medical Research* 19.1, pp. 71–99. DOI: 10.1177/0962280209105020.

Andersen, P. K. et al. (June 2012). "Competing risks in epidemiology: possibilities and pitfalls". In: *International Journal of Epidemiology* 41.3, pp. 861–870. DOI: 10.1093/ije/dyr213.

Andrinopoulou, E.-R. et al. (Aug. 2017). "Combined dynamic predictions using joint models of two longitudinal outcomes and competing risk data". en. In: *Statistical Methods in Medical Research* 26.4, pp. 1787–1801. DOI: 10.1177/0962280215588340.

Atassi, N. et al. (Nov. 2014). "The PRO-ACT database: design, initial analyses, and predictive features". eng. In: *Neurology* 83.19, pp. 1719–1725. DOI: 10.1212/WNL.0000000000000951.

Balestrino, R. and A. H. V. Schapira (Jan. 2020). "Parkinson disease". eng. In: *European Journal of Neurology* 27.1, pp. 27–42. DOI: 10.1111/ene.14108.

Beaulieu-Jones, B. K., C. S. Greene, and Pooled Resource Open-Access ALS Clinical Trials Consortium (Dec. 2016). "Semi-supervised learning of the electronic health record for phenotype stratification". eng. In: *Journal of Biomedical Informatics* 64, pp. 168–178. DOI: 10.1016/j.jbi.2016.10.007.

Bede, P. et al. (Aug. 2022). "Clusters of anatomical disease-burden patterns in ALS: a data-driven approach confirms radiological subtypes". en. In: *Journal of Neurology* 269.8, pp. 4404–4413. DOI: 10.1007/s00415-022-11081-3.

Bedre, R. (2020). *reneshbedre/bioinfokit: Bioinformatics data analysis and visualization toolkit.*

Beghi, E. et al. (2007). "The heterogeneity of amyotrophic lateral sclerosis: a possible explanation of treatment failure". eng. In: *Current Medicinal Chemistry* 14.30, pp. 3185–3200. DOI: 10.2174/092986707782793862.

Bensimon, G., L. Lacomblez, and V. Meininger (Mar. 1994). "A controlled trial of riluzole in amyotrophic lateral sclerosis. ALS/Riluzole Study Group". eng. In: *The New England Journal of Medicine* 330.9, pp. 585–591. DOI: 10.1056/NEJM199403033300901.

Betancourt, M. (July 2018). *A Conceptual Introduction to Hamiltonian Monte Carlo*. en.

Bie, R. M. A. d. et al. (May 2020). "Initiation of pharmacological therapy in Parkinson's disease: when, why, and how". English. In: *Lancet neurology* 19.5, pp. 452–461. DOI: 10.1016/S1474-4422(20)30036-3.

Biganzoli, E. M. et al. (June 2006). "Artificial neural network for the joint modelling of discrete cause-specific hazards". In: *Artificial Intelligence in Medicine* 37.2, pp. 119–130. DOI: 10.1016/j.artmed.2006.01.004.

Bilgel, M. and B. M. Jedynak (Dec. 2019). "Predicting time to dementia using a quantitative template of disease progression". en. In: *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 11, pp. 205–215. DOI: 10.1016/j.dadm.2019.01.005.

Bilgel, M. et al. (July 2016). "A multivariate nonlinear mixed effects model for longitudinal image analysis: Application to amyloid imaging". eng. In: *NeuroImage* 134, pp. 658–670. DOI: 10.1016/j.neuroimage.2016.04.001.

Blaak, E. (Nov. 2001). "Gender differences in fat metabolism". eng. In: *Current Opinion in Clinical Nutrition and Metabolic Care* 4.6, pp. 499–502. DOI: 10.1097/00075197-200111000-00006.

Blanche, P., M. W. Kattan, and T. A. Gerds (Apr. 2019). "The c-index is not proper for the evaluation of t-year predicted risks". en. In: *Biostatistics* 20.2, pp. 347–357. DOI: 10.1093/biostatistics/kxy006.

Blanche, P. et al. (Mar. 2015). "Quantifying and Comparing Dynamic Predictive Accuracy of Joint Models for Longitudinal Marker and Time-to-Event in Presence of Censoring and Competing Risks". en. In: *Biometrics* 71.1, pp. 102–113. DOI: 10.1111/biom.12232.

Bouaziz, O. (n.d.). "The effect of ignoring censoring in survival analysis: theoretical and practical considerations". en. In: ().

Bourke, S. C. et al. (Feb. 2006). "Effects of non-invasive ventilation on survival and quality of life in patients with amyotrophic lateral sclerosis: a randomised controlled trial". English. In: *The Lancet Neurology* 5.2, pp. 140–147. DOI: 10.1016/S1474-4422(05)70326-4.

Box, G. E. P. and D. R. Cox (1964). "An Analysis of Transformations". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 26.2, pp. 211–252.

Box, G. E. P. (1979). "Robustness in the Strategy of Scientific Model Building". en. In: *Robustness in Statistics*. Elsevier, pp. 201–236. ISBN: 978-0-12-438150-6. DOI: 10.1016/B978-0-12-438150-6.50018-2.

Breiman, L. (Oct. 2001). "Random Forests". en. In: *Machine Learning* 45.1, pp. 5–32. DOI: 10.1023/A:1010933404324.

Brooks, B. R. (Oct. 1996). "Natural history of ALS: symptoms, strength, pulmonary function, and disability". eng. In: *Neurology* 47.4 Suppl 2, S71–81; discussion S81–82. DOI: 10.1212/wnl.47.4_suppl_2.71s.

Brown, R. H. and A. Al-Chalabi (July 2017). "Amyotrophic Lateral Sclerosis". eng. In: *The New England Journal of Medicine* 377.2, pp. 162–172. DOI: 10.1056/NEJMra1603471.

Brändli, O. et al. (Mar. 1996). "Lung function in healthy never smoking adults: reference values and lower limits of normal of a Swiss population". eng. In: *Thorax* 51.3, pp. 277–283. DOI: 10.1136/thx.51.3.277.

Bull, L. M. et al. (July 2020). "Harnessing repeated measurements of predictor variables for clinical risk prediction: a review of existing methods". en. In: *Diagnostic and Prognostic Research* 4.1, p. 9. DOI: 10.1186/s41512-020-00078-z.

Casella, G. and E. I. George (1992). "Explaining the Gibbs Sampler". en. In: 46.3, pp. 167–174.

Cheng, S. C., J. P. Fine, and L. J. Wei (Mar. 1998). "Prediction of cumulative incidence function under the proportional hazards model". eng. In: *Biometrics* 54.1, pp. 219–228.

Chio, A. et al. (2009). "Prognostic factors in ALS: A critical review". In: *Amyotrophic lateral sclerosis : official publication of the World Federation of Neurology Research Group on Motor Neuron Diseases* 10.5-6, pp. 310–323. DOI: 10.3109/17482960802566824.

Chiò, A. et al. (Feb. 2020). "ALS phenotype is influenced by age, sex, and genetics: A population-based study". eng. In: *Neurology* 94.8, e802–e810. DOI: 10.1212/WNL.0000000000008869.

Christodoulou, E. et al. (June 2019). "A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models". eng. In: *Journal of Clinical Epidemiology* 110, pp. 12–22. DOI: 10.1016/j.jclinepi.2019.02.004.

Couronne, R. et al. (Apr. 2019). "Learning Disease Progression Models With Longitudinal Data and Missing Values". en. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. Venice, Italy: IEEE, pp. 1033–1037. ISBN: 978-1-5386-3641-1. DOI: 10.1109/ISBI.2019.8759198.

Cox, D. R. (1972). "Regression Models and Life-Tables". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 34.2, pp. 187–220.

Cummings, J. (Mar. 2018). "Lessons Learned from Alzheimer Disease: Clinical Trials with Negative Outcomes". In: *Clinical and Translational Science* 11.2, pp. 147–152. DOI: 10.1111/cts.12491.

Daghlas, S. A., R. Govindarajan, and Pooled Resource Open-Access ALS Clinical Trials Consortium (Sept. 2021). "Relative effects of forced vital capacity and ALSFRS-R on survival in ALS". eng. In: *Muscle & Nerve* 64.3, pp. 346–351. DOI: 10.1002/mus.27344.

Davidson-Pilon, C. (May 2023). *lifelines, survival analysis in Python*. DOI: 10.5281/zenodo.7883870.

Daxberger, E. et al. (2021). "Laplace Redux - Effortless Bayesian Deep Learning". In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., pp. 20089–20103.

Degens, P. and R. Merget (Mar. 2008). "Reference values for spirometry of the European Coal and Steel Community: time for change". en. In: *European Respiratory Journal* 31.3, pp. 687–688. DOI: 10.1183/09031936.00145507.

Devaux, A. et al. (Feb. 2023). *Random survival forests with multivariate longitudinal endogenous covariates*. DOI: 10.48550/arXiv.2208.05801.

Di Folco, C. et al. (2023). "Charting Disease Trajectories from Isolated REM Sleep Behavior Disorder to Parkinson's Disease". en. In: *Movement Disorders* n/a.n/a. DOI: 10.1002/mds.29662.

Dibling, M. et al. (2024). "Care pathway heterogeneity in Amyotrophic Lateral Sclerosis: effects of gender, age and onset." en. In: *[Manuscript submitted for publication]*.

Donohue, M. C. et al. (Oct. 2014). "Estimating long-term multivariate progression from short-term data". In: *Alzheimer's & dementia : the journal of the Alzheimer's Association* 10.0, S400–S410. DOI: 10.1016/j.jalz.2013.10.003.

Duara, R. and W. Barker (Jan. 2022). "Heterogeneity in Alzheimer's Disease Diagnosis and Progression Rates: Implications for Therapeutic Trials". en. In: *Neurotherapeutics* 19.1, pp. 8–25. DOI: 10.1007/s13311-022-01185-z.

Dugué, A. E. et al. (Nov. 2016). "How to Deal with Interval-Censored Data Practically while Assessing the Progression-Free Survival: A Step-by-Step Guide Using SAS and R Software". In: *Clinical Cancer Research* 22.23, pp. 5629–5635. DOI: 10.1158/1078-0432.CCR-16-1017.

Dupuis, L. et al. (Jan. 2011). "Energy metabolism in amyotrophic lateral sclerosis". eng. In: *The Lancet. Neurology* 10.1, pp. 75–82. DOI: 10.1016/S1474-4422(10)70224-6.

Durrleman, S. et al. (May 2013). "Toward a comprehensive framework for the spatiotemporal statistical analysis of longitudinal shape data". eng. In: *International Journal of Computer Vision* 103.1, pp. 22–59. DOI: 10.1007/s11263-012-0592-x.

Eijk, R. P. A. van et al. (Sept. 2019). "Two heads are better than one: benefits of joint models for ALS trials". eng. In: *Journal of Neurology, Neurosurgery, and Psychiatry* 90.9, pp. 1071–1072. DOI: 10.1136/jnnp-2019-320553.

Elamin, M. et al. (June 2015). "Predicting prognosis in amyotrophic lateral sclerosis: a simple algorithm". eng. In: *Journal of Neurology* 262.6, pp. 1447–1454. DOI: 10.1007/s00415-015-7731-6.

Elashoff, R. M., G. Li, and N. Li (June 2007). "An approach to joint analysis of longitudinal measurements and competing risks failure time data". eng. In: *Statistics in Medicine* 26.14, pp. 2813–2835. DOI: 10.1002/sim.2749.

— (Aug. 2016). *Joint Modeling of Longitudinal and Time-to-event Data*. en. CRC Press, Taylor & Francis Group. ISBN: 978-1-4398-0782-8.

Etzkorn, L. H. et al. (2024). "A joint frailty model for recurrent and competing terminal events: Application to delirium in the ICU". en. In: *Statistics in Medicine* n/a.n/a. DOI: 10.1002/sim.10053.

Fang, T. et al. (Apr. 2017). "Comparison of the King's and MiToS staging systems for ALS". In: *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration* 18.3-4, pp. 227–232. DOI: 10.1080/21678421.2016.1265565.

Farrington, D. P. (May 1991). "Longitudinal Research Strategies: Advantages, Problems, and Prospects". en. In: *Journal of the American Academy of Child & Adolescent Psychiatry* 30.3, pp. 369–374. DOI: 10.1097/00004583-199105000-00003.

FDA (2019). "Amyotrophic Lateral Sclerosis: Developing Drugs for Treatment Guidance for Industry". en. In: p. 11.

— (2023). "Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products". en. In.

Feeney, J. et al. (May 2016). "Measurement Error, Reliability, and Minimum Detectable Change in the Mini-Mental State Examination, Montreal Cognitive Assessment, and Color Trails Test among Community Living Middle-Aged and Older Adults". eng. In: *Journal of Alzheimer's disease: JAD* 53.3, pp. 1107–1114. DOI: 10.3233/JAD-160248.

Ferrer, L. et al. (2016). "Joint modelling of longitudinal and multi-state processes: application to clinical progressions in prostate cancer". en. In: *Statistics in Medicine* 35.22, pp. 3933–3948. DOI: 10.1002/sim.6972.

Fine, J. P. and R. J. Gray (1999). "A Proportional Hazards Model for the Subdistribution of a Competing Risk". In: *Journal of the American Statistical Association* 94.446, pp. 496–509. DOI: 10.2307/2670170.

Fleming, T. R. and D. P. Harrington (Aug. 2013). *Counting Processes and Survival Analysis*. en. John Wiley & Sons. ISBN: 978-0-471-76988-0.

Fournier, C. N., V. James, and J. D. Glass (Apr. 2023). "Clinically meaningful change: evaluation of the Rasch-built Overall Amyotrophic Lateral Sclerosis Disability Scale (ROADS) and the ALSFRS-R". In: *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* 24.3-4, pp. 311–316. DOI: 10.1080/21678421.2022.2153607.

Fournier, N. and S. Durrleman (2023). "A Multimodal Disease Progression Model for Genetic Associations with Disease Dynamics". en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*. Ed. by H. Greenspan et al. Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, pp. 601–610. ISBN: 978-3-031-43904-9. DOI: 10.1007/978-3-031-43904-9_58.

Furgal, A. K., A. Sen, and J. M. Taylor (2019). "Review and Comparison of Computational Approaches for Joint Longitudinal and Time-to-Event Models". en. In: *International Statistical Review* 87.2, pp. 393–418. DOI: 10.1111/insr.12322.

Georges, M. et al. (Feb. 2017). "Initiation of non-invasive ventilation in amyotrophic lateral sclerosis and clinical practice guidelines: Single-centre, retrospective, descriptive study in a national reference centre". eng. In: *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration* 18.1-2, pp. 46–52. DOI: 10.1080/21678421.2016.1236817.

Georgoulopoulou, E. et al. (Sept. 2013). "The impact of clinical factors, riluzole and therapeutic interventions on ALS survival: A population based study in Modena, Italy". en. In: *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* 14.5-6, pp. 338–345. DOI: 10.3109/21678421.2013.763281.

Gibbons, R. D., D. Hedeker, and S. DuToit (Apr. 2010). "Advances in Analysis of Longitudinal Data". In: *Annual review of clinical psychology* 6, pp. 79–107. DOI: 10.1146/annurev.clinpsy.032408.153550.

Goetz, C. G. et al. (Nov. 2008). "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results". eng. In: *Movement Disorders: Official Journal of the Movement Disorder Society* 23.15, pp. 2129–2170. DOI: 10.1002/mds.22340.

Gordon, P. H. et al. (Oct. 2010). "Progression in ALS is not linear but is curvilinear". eng. In: *Journal of Neurology* 257.10, pp. 1713–1717. DOI: 10.1007/s00415-010-5609-1.

Gorfine, M. and L. Hsu (June 2011). "Frailty-Based Competing Risks Model for Multivariate Survival Data". In: *Biometrics* 67.2, pp. 415–426. DOI: 10.1111/j.1541-0420.2010.01470.x.

Grassano, M. et al. (2024). "Sex Differences in Amyotrophic Lateral Sclerosis Survival and Progression: A Multidimensional Analysis". en. In: *Annals of Neurology* n/a.n/a. DOI: 10.1002/ana.26933.

Greenland, J. C., C. H. Williams-Gray, and R. A. Barker (2019). "The clinical heterogeneity of Parkinson's disease and its therapeutic implications". en. In: *European Journal of Neuroscience* 49.3, pp. 328–338. DOI: 10.1111/ejn.14094.

Grollemund, V. et al. (Feb. 2019). "Machine Learning in Amyotrophic Lateral Sclerosis: Achievements, Pitfalls, and Future Directions". en. In: *Frontiers in Neuroscience* 13, p. 135. DOI: 10.3389/fnins.2019.00135.

Guillot, S. J., M. Bolborea, and L. Dupuis (Oct. 2021). "Dysregulation of energy homeostasis in amyotrophic lateral sclerosis". eng. In: *Current Opinion in Neurology* 34.5, pp. 773–780. DOI: 10.1097/WCO.0000000000000982.

Hao, L. et al. (2024). "Joint AFT random-effect modeling approach for clustered competing-risks data". In: *Journal of Statistical Computation and Simulation* 0.0, pp. 1–29. DOI: 10.1080/00949655.2024.2319188.

Haverkamp, L. J., V. Appel, and S. H. Appel (June 1995). "Natural history of amyotrophic lateral sclerosis in a database population. Validation of a scoring system and a model for survival prediction". eng. In: *Brain: A Journal of Neurology* 118 (Pt 3), pp. 707–719. DOI: 10.1093/brain/118.3.707.

Hesters, A. et al. (Sept. 2020). "Predictive factors for prognosis after gastrostomy placement in routine non-invasive ventilation users ALS patients". en. In: *Scientific Reports* 10.1, p. 15117. DOI: 10.1038/s41598-020-70422-2.

Hickey, G. L. et al. (Sept. 2016). "Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues". In: *BMC Medical Research Methodology* 16.1, p. 117. DOI: 10.1186/s12874-016-0212-5.

Hirose, T. et al. (Dec. 2018). "Clinical characteristics of long-term survival with noninvasive ventilation and factors affecting the transition to invasive ventilation in amyotrophic lateral sclerosis". eng. In: *Muscle & Nerve* 58.6, pp. 770–776. DOI: 10.1002/mus.26149.

Hothorn, T. and H. H. Jung (Sept. 2014). "RandomForest4Life: a Random Forest for predicting ALS disease progression". eng. In: *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration* 15.5-6, pp. 444–452. DOI: 10.3109/21678421.2014.893361.

Hougaard, P. (1995). "Frailty models for survival data". eng. In: *Lifetime Data Analysis* 1.3, pp. 255–273. DOI: 10.1007/BF00985760.

Hoffman, M. D. and A. Gelman (2014). "The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo". en. In.

Huang, X., G. Li, and R. M. Elashoff (2010). "A joint model of longitudinal and competing risks survival data with heterogeneous random effects and outlying longitudinal measurements". In: *Statistics and its interface* 3.2, pp. 185–195.

Huang, Z. et al. (Dec. 2017). "Complete hazard ranking to analyze right-censored data: An ALS survival study". eng. In: *PLoS computational biology* 13.12, e1005887. DOI: 10.1371/journal.pcbi.1005887.

Iddi, S. et al. (2018). "Estimating the evolution of disease in the Parkinson's Progression Markers Initiative". In: *Neuro-degenerative diseases* 18.4, pp. 173–190. DOI: 10.1159/000488780.

Ishwaran, H. et al. (Feb. 2011). "Random survival forests for high-dimensional data". en. In: *Statistical Analysis and Data Mining: The ASA Data Science Journal* 4.1, pp. 115–132. DOI: 10.1002/sam.10103.

Ishwaran, H. et al. (Oct. 2014). "Random survival forests for competing risks". In: *Biostatistics* 15.4, pp. 757–773. DOI: 10.1093/biostatistics/kxu010.

Jablecki, C. K., C. Berry, and J. Leach (Oct. 1989). "Survival prediction in amyotrophic lateral sclerosis". eng. In: *Muscle & Nerve* 12.10, pp. 833–841. DOI: 10.1002/mus.880121008.

Jack, C. R. et al. (Jan. 2010). "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade". eng. In: *The Lancet. Neurology* 9.1, pp. 119–128. DOI: 10.1016/S1474-4422(09)70299-6.

Jack, C. R. et al. (Mar. 2013). "Brain β-amyloid load approaches a plateau". eng. In: *Neurology* 80.10, pp. 890–896. DOI: 10.1212/WNL.0b013e3182840bbe.

Jackson, C. (2016). "**flexsurv** : A Platform for Parametric Survival Modeling in *R*". en. In: *Journal of Statistical Software* 70.8. DOI: 10.18637/jss.v070.i08.

Jackson, D. et al. (Nov. 2014). "Relaxing the independent censoring assumption in the Cox proportional hazards model using multiple imputation". In: *Statistics in Medicine* 33.27, pp. 4681–4694. DOI: 10.1002/sim.6274.

Jahandideh, S. et al. (Apr. 2018). "Longitudinal modeling to predict vital capacity in amyotrophic lateral sclerosis". en. In: *Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration* 19.3-4, pp. 294–302. DOI: 10.1080/21678421.2017.1418003.

Jarrett, D., J. Yoon, and M. Schaar (July 2019). "Dynamic Prediction in Clinical Survival Analysis Using Temporal Convolutional Networks". In: *IEEE Journal of Biomedical and Health Informatics* PP, pp. 1–1. DOI: 10.1109/JBHI.2019.2929264.

Jedynak, B. M. et al. (Nov. 2012). "A Computational Neurodegenerative Disease Progression Score: Method and Results with the Alzheimer's Disease Neuroimaging Initiative Cohort". In: *NeuroImage* 63.3, pp. 1478–1486. DOI: 10.1016/j.neuroimage.2012.07.059.

Jiang, R. and D. N. P. Murthy (Dec. 2011). "A study of Weibull shape parameter: Properties and significance". In: *Reliability Engineering & System Safety* 96.12, pp. 1619–1626. DOI: 10.1016/j.ress.2011.09.003.

Jiang, S., Y. Xie, and G. Colditz (Feb. 2020). *Functional Ensemble Survival Tree: Dynamic Prediction of Alzheimer's Disease Progression Accommodating Multiple Time-Varying Covariates*. DOI: 10.1101/2020.02.17.952994.

Kalbfleisch, J. D. and R. L. Prentice (Sept. 2002). *The Statistical Analysis of Failure Time Data*. en. Wiley. ISBN: 978-0-471-36357-6.

Kantidakis, G. et al. (Feb. 2023). "Statistical models versus machine learning for competing risks: development and validation of prognostic models". In: *BMC Medical Research Methodology* 23.1, p. 51. DOI: 10.1186/s12874-023-01866-z.

Kaplan, E. L. and P. Meier (June 1958). "Nonparametric Estimation from Incomplete Observations". In: *Journal of the American Statistical Association* 53.282, pp. 457–481. DOI: 10.1080/01621459.1958.10501452.

Kim, S. et al. (Nov. 2012). "Joint Models of Longitudinal Data and Recurrent Events with Informative Terminal Event". In: *Statistics in biosciences* 4.2, pp. 262–281. DOI: 10.1007/s12561-012-9061-x.

Kleopa, K. A. et al. (Mar. 1999). "Bipap improves survival and rate of pulmonary function decline in patients with ALS". English. In: *Journal of the Neurological Sciences* 164.1, pp. 82–88. DOI: 10.1016/S0022-510X(99)00045-3.

Ko, K. D. et al. (2014). "Predicting the severity of motor neuron disease progression using electronic health record data with a cloud computing Big Data approach". en. In: *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology proceedings IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology* 2014. DOI: 10.1109/CIBCB.2014.6845506.

Koval, I. (2020). "Learning Multimodal Digital Models of Disease Progression from Longitudinal Data: Methods & Algorithms for the Description, Prediction and Simulation of Alzheimer's Disease Progression". en. In: p. 181.

Koval, I. et al. (Apr. 2021). "AD Course Map charts Alzheimer's disease progression". eng. In: *Scientific Reports* 11.1, p. 8020. DOI: 10.1038/s41598-021-87434-1.

Koval, I. et al. (2022). "Forecasting individual progression trajectories in Huntington disease enables more powered clinical trials". en. In: *Scientific Reports* 12.1, p. 18928. DOI: 10.1038/s41598-022-18848-8.

Kueffner, R. et al. (Jan. 2019). "Stratification of amyotrophic lateral sclerosis patients: a crowdsourcing approach". In: *Scientific Reports* 9, p. 690. DOI: 10.1038/s41598-018-36873-4.

Kuhn, E. and M. Lavielle (Aug. 2004). "Coupling a stochastic approximation version of EM with an MCMC procedure". en. In: *ESAIM: Probability and Statistics* 8, pp. 115–131. DOI: 10.1051/ps:2004007.

Kurland, B. F. et al. (2009). "Longitudinal Data with Follow-up Truncated by Death: Match the Analysis Method to Research Aims". In: *Statistical science : a review journal of the Institute of Mathematical Statistics* 24.2, p. 211. DOI: 10.1214/09-STS293.

Kyheng, M. et al. (Sept. 2021). "Joint latent class model: Simulation study of model properties and application to amyotrophic lateral sclerosis disease". eng. In: *BMC medical research methodology* 21.1, p. 198. DOI: 10.1186/s12874-021-01377-9.

Körner, S. et al. (July 2013). "Weight loss, dysphagia and supplement intake in patients with amyotrophic lateral sclerosis (ALS): impact on quality of life and therapeutic options". eng. In: *BMC neurology* 13, p. 84. DOI: 10.1186/1471-2377-13-84.

Küffner, R. et al. (Jan. 2015). "Crowdsourced analysis of clinical trial data to predict amyotrophic lateral sclerosis progression". eng. In: *Nature Biotechnology* 33.1, pp. 51–57. DOI: 10.1038/nbt.3051.

Lahouel, K. et al. (Nov. 2023). *Learning nonparametric ordinary differential equations from noisy data*. DOI: 10.48550/arXiv.2206.15215.

Laird, N. M. and J. H. Ware (1982). "Random-Effects Models for Longitudinal Data". In: *Biometrics* 38.4, pp. 963–974. DOI: 10.2307/2529876.

Landau, S. et al. (2021). "Florbetapir (AV45) processing methods". en. In.

Lavalley-Morelle, A. et al. (Apr. 2024). "Extending the code in the open-source saemix package to fit joint models of longitudinal and time-to-event data". en. In: *Computer Methods and Programs in Biomedicine* 247, p. 108095. DOI: 10.1016/j.cmpb.2024.108095.

Lavielle, M. and L. Aarons (Feb. 2016). "What do we mean by identifiability in mixed effects models?" en. In: *Journal of Pharmacokinetics and Pharmacodynamics* 43.1, pp. 111–122. DOI: 10.1007/s10928-015-9459-4.

Lay, K. and G. Li (Sept. 2019). "Do microglial sex differences contribute to sex differences in neurodegenerative diseases?" In: *Trends in molecular medicine* 25.9, pp. 741–749. DOI: 10.1016/j.molmed.2019.05.001.

Lee, C., J. Yoon, and M. V. D. Schaar (Jan. 2020). "Dynamic-DeepHit: A Deep Learning Approach for Dynamic Survival Analysis With Competing Risks Based on Longitudinal Data". en. In: *IEEE Transactions on Biomedical Engineering* 67.1, pp. 122–133. DOI: 10.1109/TBME.2019.2909027.

Lee, C. et al. (Apr. 2018). "DeepHit: A Deep Learning Approach to Survival Analysis With Competing Risks". en. In: *Proceedings of the AAAI Conference on Artificial Intelligence* 32.1. DOI: 10.1609/aaai.v32i1.11842.

Lee, H. et al. (Dec. 2021). "Multi-omic analysis of selectively vulnerable motor neuron subtypes implicates altered lipid metabolism in ALS". en. In: *Nature Neuroscience* 24.12, pp. 1673–1685. DOI: 10.1038/s41593-021-00944-z.

Leffondré, K. et al. (Aug. 2013). "Interval-censored time-to-event and competing risk with death: is the illness-death model more accurate than the Cox model?" eng. In: *International Journal of Epidemiology* 42.4, pp. 1177–1186. DOI: 10.1093/ije/dyt126.

Leung, K.-M., R. M. Elashoff, and A. A. Afifi (May 1997). "CENSORING ISSUES IN SURVIVAL ANALYSIS". en. In: *Annual Review of Public Health* 18.1, pp. 83–104. DOI: 10.1146/annurev.publhealth.18.1.83.

Li, D. et al. (Mar. 2017). "Bayesian latent time joint mixed effect models for multicohort longitudinal data". en. In: *Statistical Methods in Medical Research* 28.3, pp. 835–845. DOI: 10.1177/0962280217737566.

Li, J.-Y. et al. (Aug. 2022). "Correlation of weight and body composition with disease progression rate in patients with amyotrophic lateral sclerosis". eng. In: *Scientific Reports* 12.1, p. 13292. DOI: 10.1038/s41598-022-16229-9.

Liang, K.-Y. and S. L. Zeger (Apr. 1986). "Longitudinal data analysis using generalized linear models". In: *Biometrika* 73.1, pp. 13–22. DOI: 10.1093/biomet/73.1.13.

Lin, H. et al. (Mar. 2002). "Latent Class Models for Joint Analysis of Longitudinal Biomarker and Event Process Data: Application to Longitudinal Prostate-Specific Antigen Readings and Prostate Cancer". en. In: *Journal of the American Statistical Association* 97.457, pp. 53–65. DOI: 10.1198/016214502753479220.

Lin, J., K. Li, and S. Luo (Jan. 2021). "Functional survival forests for multivariate longitudinal outcomes: Dynamic prediction of Alzheimer's disease progression". eng. In: *Statistical Methods in Medical Research* 30.1, pp. 99–111. DOI: 10.1177/0962280220941532.

Liu, L. C. and D. Hedeker (Mar. 2006). "A mixed-effects regression model for longitudinal multivariate ordinal data". eng. In: *Biometrics* 62.1, pp. 261–268. DOI: 10.1111/j.1541-0420.2005.00408.x.

Lo Coco, D. et al. (Sept. 2006). "Noninvasive positive-pressure ventilation in ALS". In: *Neurology* 67.5, pp. 761–765. DOI: 10.1212/01.wnl.0000227785.73714.64.

Locascio, J. J. and A. Atri (2011). "An Overview of Longitudinal Data Analysis Methods for Neurological Research". In: *Dementia and Geriatric Cognitive Disorders Extra* 1.1, pp. 330–357. DOI: 10.1159/000330228.

Locke, S. et al. (June 2021). "Natural language processing in medicine: A review". en. In: *Trends in Anaesthesia and Critical Care* 38, pp. 4–9. DOI: 10.1016/j.tacc.2021.02.007.

Logroscino, G. et al. (Apr. 2010). "Incidence of Amyotrophic Lateral Sclerosis in Europe". In: *Journal of neurology, neurosurgery, and psychiatry* 81.4, pp. 385–390. DOI: 10.1136/jnnp.2009.183525.

Lorenzi, M. et al. (Apr. 2019). "Probabilistic disease progression modeling to characterize diagnostic uncertainty: Application to staging and prediction in Alzheimer's disease". eng. In: *NeuroImage* 190, pp. 56–68. DOI: 10.1016/j.neuroimage.2017.08.059.

Lu, X. et al. (2023). "A two-level copula joint model for joint analysis of longitudinal and competing risks data". en. In: *Statistics in Medicine* 42.12, pp. 1909–1930. DOI: 10.1002/sim.9704.

Maheux, E. et al. (2023). "Forecasting individual progression trajectories in Alzheimer's disease". en. In: *Nature Communications* 14.1, p. 761. DOI: 10.1038/s41467-022-35712-5.

Marinescu, R. V. et al. (Oct. 2019). "TADPOLE Challenge: Accurate Alzheimer's disease prediction through crowdsourced forecasting of future data". eng. In: *PRedictive Intelligence in MEdicine. PRIME (Workshop)* 11843, pp. 1–10. DOI: 10.1007/978-3-030-32281-6_1.

Masci, C., M. Spreafico, and F. Ieva (Nov. 2023). *Joint modelling of recurrent and terminal events with discretely-distributed non-parametric frailty: application on re-hospitalizations and death in heart failure patients.*

Mauguen, A. et al. (2013). "Dynamic prediction of risk of death using history of cancer recurrences in joint frailty models". en. In: *Statistics in Medicine* 32.30, pp. 5366–5380. DOI: 10.1002/sim.5980.

McCulloch, C. E., S. R. Searle, and J. M. Neuhaus (June 2008). *Generalized, Linear, and Mixed Models*. en. Wiley. ISBN: 978-0-470-07371-1.

McKenna, M et al. (1998). "Chronic disease epidemiology and control". In.

McNealis, V., E. E. M. Moodie, and N. Dean (Apr. 2024). *Joint mixed-effects models for causal inference in clustered network-based observational studies*. en.

Mejia, A. F. et al. (July 2022). "Longitudinal surface-based spatial Bayesian GLM reveals complex trajectories of motor neurodegeneration in ALS". eng. In: *NeuroImage* 255, p. 119180. DOI: 10.1016/j.neuroimage.2022.119180.

Miano, B. et al. (Dec. 2004). "Inter-evaluator reliability of the ALS functional rating scale". eng. In: *Amyotrophic Lateral Sclerosis and Other Motor Neuron Disorders: Official Publication of the World Federation of Neurology, Research Group on Motor Neuron Diseases* 5.4, pp. 235–239. DOI: 10.1080/14660820410021302.

Millar, R. B. (Mar. 2018). "Conditional vs marginal estimation of the predictive loss of hierarchical models using WAIC and cross-validation". en. In: *Statistics and Computing* 28.2, pp. 375–385. DOI: 10.1007/s11222-017-9736-8.

Miller, R. G. et al. (Jan. 2007). "Riluzole for amyotrophic lateral sclerosis (ALS)/motor neuron disease (MND)". eng. In: *The Cochrane Database of Systematic Reviews* 1, p. CD001447. DOI: 10.1002/14651858.CD001447.pub2.

Moglia, C. et al. (June 2019). "Early weight loss in amyotrophic lateral sclerosis: outcome relevance and clinical correlates in a population-based cohort". eng. In: *Journal of Neurology, Neurosurgery, and Psychiatry* 90.6, pp. 666–673. DOI: 10.1136/jnnp-2018-319611.

Morelot-Panzini, C., G. Bruneteau, and J. Gonzalez-Bermejo (2019). "NIV in amyotrophic lateral sclerosis: The 'when' and 'how' of the matter". en. In: *Respirology* 24.6, pp. 521–530. DOI: 10.1111/resp.13525.

Moulaire, P. et al. (2023). "Temporal Dynamics of the Scale for the Assessment and Rating of Ataxia in Spinocerebellar Ataxias". en. In: *Movement Disorders* 38.1, pp. 35–44. DOI: 10.1002/mds.29255.

Mozumder, S. I., M. J. Rutherford, and P. C. Lambert (Jan. 2018). "Direct likelihood inference on the cause-specific cumulative incidence function: a flexible parametric regression modelling approach". In: *Statistics in medicine* 37.1, pp. 82–97. DOI: 10.1002/sim.7498.

Mueller, S. G. et al. (Nov. 2005). "The Alzheimer's Disease Neuroimaging Initiative". In: *Neuroimaging Clinics of North America*. Alzheimer's Disease: 100 Years of Progress 15.4, pp. 869–877. DOI: 10.1016/j.nic.2005.09.008.

Murawska, M., D. Rizopoulos, and E. Lesaffre (Feb. 2012). "A Two-Stage Joint Model for Nonlinear Longitudinal Response and a Time-to-Event with Application in Transplantation Studies". en. In: *Journal of Probability and Statistics* 2012, e194194. DOI: 10.1155/2012/194194.

Murdock, B. J. et al. (Mar. 2021). "Amyotrophic Lateral Sclerosis Survival Associates With Neutrophils in a Sex-specific Manner". eng. In: *Neurology(R) Neuroimmunology & Neuroinflammation* 8.2, e953. DOI: 10.1212/NXI.0000000000000953.

Musoro, J. Z., R. B. Geskus, and A. H. Zwinderman (2015). "A joint model for repeated events of different types and multiple longitudinal outcomes with application to a follow-up study of patients after kidney transplant". en. In: *Biometrical Journal* 57.2, pp. 185–200. DOI: 10.1002/bimj.201300167.

Müller, H.-P. et al. (June 2016). "A large-scale multicentre cerebral diffusion tensor imaging study in amyotrophic lateral sclerosis". eng. In: *Journal of Neurology, Neurosurgery, and Psychiatry* 87.6, pp. 570–579. DOI: 10.1136/jnnp-2015-311952.

Nahhas, R. W. (2024). *9.2 MCAR, MAR, MNAR | Introduction to Regression Methods for Public Health Using R.*

Nakamura, R. et al. (Feb. 2022). "Investigation of the prognostic predictive value of serum lipid profiles in amyotrophic lateral sclerosis: roles of sex and hypermetabolism". en. In: *Scientific Reports* 12.1, p. 1826. DOI: 10.1038/s41598-022-05714-w.

Nelson, A. T. and D. Trotti (July 2022). "Altered Bioenergetics and Metabolic Homeostasis in Amyotrophic Lateral Sclerosis". eng. In: *Neurotherapeutics: The Journal of the American Society for Experimental NeuroTherapeutics* 19.4, pp. 1102–1118. DOI: 10.1007/s13311-022-01262-3.

Nesterov, Y. (2018). *Lectures on Convex Optimization*. en. Vol. 137. Springer Optimization and Its Applications. Cham: Springer International Publishing. ISBN: 978-3-319-91577-7 978-3-319-91578-4. DOI: 10.1007/978-3-319-91578-4.

Nguyen, V. T. et al. (Sept. 2023). *FLASH: a Fast joint model for Longitudinal And Survival data in High dimension*. en. DOI: arXiv:2309.03714[stat]. arXiv: 2309.03714 [stat.ME].

Nicolaie, M. A., H. C. Van Houwelingen, and H. Putter (May 2010). "Vertical modeling: A pattern mixture approach for competing risks modeling". en. In: *Statistics in Medicine* 29.11, pp. 1190–1205. DOI: 10.1002/sim.3844.

Niu, H. et al. (Oct. 2017). "Prevalence and incidence of Alzheimer's disease in Europe: A meta-analysis". eng, spa. In: *Neurologia (Barcelona, Spain)* 32.8, pp. 523–532. DOI: 10.1016/j.nrl.2016.02.016.

Nona, R. J. et al. (Oct. 2023). "Age of Onset and Length of Survival of Queensland Patients with Amyotrophic Lateral Sclerosis: Details of Subjects with Early Onset and Subjects with Long Survival". In: *Neuro-Degenerative Diseases* 22.3-4, pp. 104–121. DOI: 10.1159/000528875.

Norris, F. et al. (Aug. 1993). "Onset, natural history and outcome in idiopathic adult motor neuron disease". eng. In: *Journal of the Neurological Sciences* 118.1, pp. 48–55. DOI: 10.1016/0022-510x(93)90245-t.

Olanow, C. W., K. Kieburtz, and A. H. V. Schapira (Dec. 2008). "Why have we failed to achieve neuroprotection in Parkinson's disease?" eng. In: *Annals of Neurology* 64 Suppl 2, S101–110. DOI: 10.1002/ana.21461.

Organization, W. H. et al. (2022). "Invisible numbers: the true extent of noncommunicable diseases and what to do about them". In.

Oriol, A.-P. et al. (2023). "PyMC: A Modern and Comprehensive Probabilistic Programming Framework in Python". In: *PeerJ Computer Science* 9, e1516. DOI: 10.7717/peerj-cs.1516.

Ortholand, J. et al. (2023). "Interaction of sex and onset site on the disease trajectory of amyotrophic lateral sclerosis". en. In: *Journal of Neurology* 270.12, pp. 5903–5912. DOI: 10.1007/s00415-023-11932-7.

Ortholand, J. et al. (Jan. 2024). "Joint model with latent disease age: overcoming the need for reference time". en. In: *[Manuscript submitted for publication]*.

Oxtoby, N. P. et al. (2014). "Learning Imaging Biomarker Trajectories from Noisy Alzheimer's Disease Data Using a Bayesian Multilevel Model". en. In: *Bayesian and grAphical Models for Biomedical Imaging*. Ed. by M. J. Cardoso et al. Cham: Springer International Publishing, pp. 85–94. ISBN: 978-3-319-12289-2. DOI: 10.1007/978-3-319-12289-2_8.

Pan, S. and A. van den Hout (Mar. 2023). "Bivariate joint models for survival and change of cognitive function". en. In: *Statistical Methods in Medical Research* 32.3, pp. 474–492. DOI: 10.1177/09622802221146307.

Parkinson Progression Marker Initiative (Dec. 2011). "The Parkinson Progression Marker Initiative (PPMI)". eng. In: *Progress in Neurobiology* 95.4, pp. 629–635. DOI: 10.1016/j.pneurobio.2011.09.005.

Pedregosa, F. et al. (n.d.). "Scikit-learn: Machine Learning in Python". en. In: *MACHINE LEARNING IN PYTHON* (), p. 6.

Peter, R. S. et al. (Oct. 2017). "Life course body mass index and risk and prognosis of amyotrophic lateral sclerosis: results from the ALS registry Swabia". en. In: *European Journal of Epidemiology* 32.10, pp. 901–908. DOI: 10.1007/s10654-017-0318-z.

Petrov, D. et al. (Mar. 2017). "ALS Clinical Trials Review: 20 Years of Failure. Are We Any Closer to Registering a New Treatment?" In: *Frontiers in Aging Neuroscience* 9, p. 68. DOI: 10.3389/fnagi.2017.00068.

Pfohl, S. R. et al. (2018). "Unraveling the Complexity of Amyotrophic Lateral Sclerosis Survival Prediction". eng. In: *Frontiers in Neuroinformatics* 12, p. 36. DOI: 10.3389/fninf.2018.00036.

Pölsterl, S. (2020). "scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn". In: *Journal of Machine Learning Research* 21.212, pp. 1–6.

Poulet, P.-E. and S. Durrleman (2023a). "Modeling individual disease trajectory: application to the prediction of treatment effect". en. PhD thesis.

— (2023b). "Multivariate disease progression modeling with longitudinal ordinal data". en. In: *Statistics in Medicine* 42.18, pp. 3164–3183. DOI: 10.1002/sim.9770.

Prentice, R. L. and L. A. Gloeckler (Mar. 1978). "Regression Analysis of Grouped Survival Data with Application to Breast Cancer Data". In: *Biometrics* 34.1, p. 57. DOI: 10.2307/2529588.

Proust-Lima, C., J.-F. Dartigues, and H. Jacqmin-Gadda (2016). "Joint modeling of repeated multivariate cognitive measures and competing risks of dementia and death: a latent process and latent class approach". en. In: *Statistics in Medicine* 35.3, pp. 382–398. DOI: 10.1002/sim.6731.

Proust-Lima, C., V. Philipps, and B. Liquet (June 2017). "Estimation of Extended Mixed Models Using Latent Classes and Latent Processes: The R Package lcmm". en. In: *Journal of Statistical Software* 78, pp. 1–56. DOI: 10.18637/jss.v078.i02.

Proust-Lima, C. et al. (Feb. 2009). "Joint modelling of multivariate longitudinal outcomes and a time-to-event: A nonlinear latent class approach". In: *Computational Statistics & Data Analysis* 53.4, pp. 1142–1154. DOI: 10.1016/j.csda.2008.10.017.

Proust-Lima, C. et al. (Feb. 2014). "Joint latent class models for longitudinal and time-to-event data: A review". en. In: *Statistical Methods in Medical Research* 23.1, pp. 74–90. DOI: 10.1177/0962280212445839.

Proust-Lima, C. et al. (2023). "Describing complex disease progression using joint latent class models for multivariate longitudinal markers and clinical endpoints". en. In: *Statistics in Medicine* n/a.n/a. DOI: 10.1002/sim.9844.

Raket, L. L. (2020). "Statistical Disease Progression Modeling in Alzheimer Disease". In: *Frontiers in Big Data* 3.

— (2022). "Progression models for repeated measures: Estimating novel treatment effects in progressive diseases". en. In: *Statistics in Medicine* 41.28, pp. 5537–5557. DOI: 10.1002/sim.9581.

Raymond, J. et al. (2021). "Reproductive History and Age of Onset for Women Diagnosed with Amyotrophic Lateral Sclerosis: Data from the National ALS Registry: 2010-2018". eng. In: *Neuroepidemiology* 55.5, pp. 416–424. DOI: 10.1159/000516344.

Reniers, W. et al. (Aug. 2017). "Prognostic value of clinical and electrodiagnostic parameters at time of diagnosis in patients with amyotrophic lateral sclerosis". eng. In: *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration* 18.5-6, pp. 341–350. DOI: 10.1080/21678421.2017.1288254.

Rizopoulos, D. (June 2012). *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. New York: Chapman and Hall/CRC. ISBN: 978-0-429-06338-1. DOI: 10.1201/b12208.

— (2016). "The *R* Package **JMbayes** for Fitting Joint Models for Longitudinal and Time-to-Event Data Using MCMC". en. In: *Journal of Statistical Software* 72.7. DOI: 10.18637/jss.v072.i07.

Rizopoulos, D. and P. Ghosh (2011). "A Bayesian semiparametric multivariate joint model for multiple longitudinal outcomes and a time-to-event". en. In: *Statistics in Medicine* 30.12, pp. 1366–1380. DOI: 10.1002/sim.4205.

Rizopoulos, D., G. Papageorgiou, and P. Miranda Afonso (2024). *JMbayes2: Extended Joint Models for Longitudinal and Time-to-Event Data*.

Robbins, H. and S. Monro (Sept. 1951). "A Stochastic Approximation Method". In: *The Annals of Mathematical Statistics* 22.3, pp. 400–407. DOI: 10.1214/aoms/1177729586.

Roblin, E., P.-H. Cournède, and S. Michiels (Jan. 2024). "Confidence intervals of survival predictions with neural networks trained on molecular data". In: *Informatics in Medicine Unlocked* 44, p. 101426. DOI: 10.1016/j.imu.2023.101426.

Rondeau, V., Y. Mazroui, and J. R. Gonzalez (2012). "**frailtypack** : An *R* Package for the Analysis of Correlated Survival Data with Frailty Models Using Penalized Likelihood Estimation or Parametrical Estimation". en. In: *Journal of Statistical Software* 47.4. DOI: 10.18637/jss.v047.i04.

Rooney, J. et al. (Sept. 2019). "The C9orf72 expansion is associated with accelerated respiratory function decline in a large Amyotrophic Lateral Sclerosis cohort". en. In: *HRB open research* 2. DOI: 10.12688/hrbopenres.12940.1.

Rooney, J. et al. (May 2017). "What does the ALSFRS-R really measure? A longitudinal and survival analysis of functional dimension subscores in amyotrophic lateral sclerosis". eng. In: *Journal of Neurology, Neurosurgery, and Psychiatry* 88.5, pp. 381–385. DOI: 10.1136/jnnp-2016-314661.

Rouanet, A. et al. (Dec. 2016). "Joint latent class model for longitudinal data and interval-censored semi-competing events: Application to dementia". eng. In: *Biometrics* 72.4, pp. 1123–1135. DOI: 10.1111/biom.12530.

Routier, A. et al. (Aug. 2021). "Clinica: An Open-Source Software Platform for Reproducible Clinical Neuroscience Studies". English. In: *Frontiers in Neuroinformatics* 15. DOI: 10.3389/fninf.2021.689675.

Royston, P. and M. K. Parmar (2002). "Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects". In: *Statistics in medicine* 21.15, pp. 2175–2197.

Rubin, D. B. (1976). "Inference and Missing Data". In: *Biometrika* 63.3, pp. 581–592. DOI: 10.2307/2335739.

Rubio, F. J. et al. (Aug. 2019). "On a general structure for hazard-based regression models: An application to population-based cancer research". en. In: *Statistical Methods in Medical Research* 28.8, pp. 2404–2417. DOI: 10.1177/0962280218782293.

Rustand, D. et al. (Feb. 2024). *Joint Modeling of Multivariate Longitudinal and Survival Outcomes with the R package INLAjoint*.

Samtani, M. N. et al. (Jan. 2013). "Disease progression model in subjects with mild cognitive impairment from the Alzheimer's disease neuroimaging initiative: CSF biomarkers predict population subtypes". In: *British Journal of Clinical Pharmacology* 75.1, pp. 146–161. DOI: 10.1111/j.1365-2125.2012.04308.x.

Saulnier, T., C. Proust-Lima, and A. Samier (Aug. 2023). *Joint analysis of disease progression markers and death using individual temporal recalibration*. ISCB44. Milan.

Saulnier, T. et al. (July 2022). "Joint models for the longitudinal analysis of measurement scales in the presence of informative dropout". In: *Methods* 203, pp. 142–151. DOI: 10.1016/j.ymeth.2022.03.003.

Sauty, B. (2023). "Multimodal modelling of Alzheimer's Disease progression". PhD thesis.

Sauty, B. and S. Durrleman (2022). "Progression Models for Imaging Data with Longitudinal Variational Auto Encoders". en. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*. Ed. by L. Wang et al. Vol. 13431. Cham: Springer Nature Switzerland, pp. 3–13. ISBN: 978-3-031-16430-9 978-3-031-16431-6. DOI: 10.1007/978-3-031-16431-6_1.

— (2023). "Impact of sex and APOE-*epsilon*4 genotype on patterns of regional brain atrophy in Alzheimer's disease and healthy aging". In: *Frontiers in Neurology* 14. DOI: https://doi.org/10.3389/fneur.2023.1161527.

Schiffman, P. L. and J. M. Belsh (Feb. 1993). "Pulmonary function at diagnosis of amyotrophic lateral sclerosis. Rate of deterioration". eng. In: *Chest* 103.2, pp. 508–513. DOI: 10.1378/chest.103.2.508.

Schiratti, J.-B. et al. (2015). "A Mixed-Effects Model with Time Reparametrization for Longitudinal Univariate Manifold-Valued Data". en. In: *Information Processing in Medical Imaging*. Ed. by S. Ourselin et al. Lecture Notes in Computer Science. Cham: Springer International Publishing, pp. 564–575. ISBN: 978-3-319-19992-4. DOI: 10.1007/978-3-319-19992-4_44.

Schiratti, J.-B. (2017). "Methods and algorithms to learn spatio-temporal changes from longitudinal manifold-valued observations". en. In: p. 200.

Schiratti, J.-B. et al. (2017). "A Bayesian Mixed-Effects Model to Learn Trajectories of Changes from Repeated Manifold-Valued Observations". en. In: p. 33.

Schluchter, M. D. (1992). "Methods for the analysis of informatively censored longitudinal data". eng. In: *Statistics in Medicine* 11.14-15, pp. 1861–1870. DOI: 10.1002/sim.4780111408.

Segura, T. et al. (Jan. 2023). "Symptoms timeline and outcomes in amyotrophic lateral sclerosis using artificial intelligence". en. In: *Scientific Reports* 13.1, p. 702. DOI: 10.1038/s41598-023-27863-2.

Shardell, M. and L. Ferrucci (Feb. 2018). "Joint mixed-effects models for causal inference with longitudinal data". In: *Statistics in medicine* 37.5, pp. 829–846. DOI: 10.1002/sim.7567.

Sherman, A. et al. (Apr. 2019). "PRO-ACE: Global Collaborative Ecosystem in a Post-PRO-ACT Era (P4.6-006)". In: *Neurology* 92.15_supplement, P4.6–006. DOI: 10.1212/WNL.92.15_supplement.P4.6-006.

Skipper, S. and J. Perktold (2010). *statsmodels: Econometric and statistical modeling with python.*

Soleimani, H., J. Hensman, and S. Saria (Aug. 2017). *Scalable Joint Models for Reliable Uncertainty-Aware Event Prediction.*

Steffen, T. and M. Seney (June 2008). "Test-retest reliability and minimal detectable change on balance and ambulation tests, the 36-item short-form health survey, and the unified Parkinson disease rating scale in people with parkinsonism". eng. In: *Physical Therapy* 88.6, pp. 733–746. DOI: 10.2522/ptj.20070214.

Sudell, M., R. Kolamunnage-Dona, and C. Tudur-Smith (Dec. 2016). "Joint models for longitudinal and time-to-event data: a review of reporting quality with a view to meta-analysis". In: *BMC Medical Research Methodology* 16, p. 168. DOI: 10.1186/s12874-016-0272-6.

Sumien, N. et al. (Nov. 2021). "Neurodegenerative Disease: Roles for Sex, Hormones, and Oxidative Stress". en. In: *Endocrinology* 162.11, bqab185. DOI: 10.1210/endocr/bqab185.

Sun, J. et al. (Nov. 2020). "Blood biomarkers and prognosis of amyotrophic lateral sclerosis". eng. In: *European Journal of Neurology* 27.11, pp. 2125–2133. DOI: 10.1111/ene.14409.

Sweeting, M. J. et al. (Dec. 2017). "The use of repeated blood pressure measures for cardiovascular risk prediction: a comparison of statistical models in the ARIC study". In: *Statistics in Medicine* 36.28, pp. 4514–4528. DOI: 10.1002/sim.7144.

Swinnen, B. and W. Robberecht (Nov. 2014). "The phenotypic variability of amyotrophic lateral sclerosis". eng. In: *Nature Reviews. Neurology* 10.11, pp. 661–670. DOI: 10.1038/nrneurol.2014.184.

Talbott, E. O., A. M. Malek, and D. Lacomis (2016). "The epidemiology of amyotrophic lateral sclerosis". eng. In: *Handbook of Clinical Neurology* 138, pp. 225–238. DOI: 10.1016/B978-0-12-802973-2.00013-6.

Tao, Y. et al. (Dec. 2021). "Trend towards virtual and hybrid conferences may be an effective climate change mitigation strategy". en. In: *Nature Communications* 12.1, p. 7324. DOI: 10.1038/s41467-021-27251-2.

Taylor, A. A. et al. (Sept. 2016). "Predicting disease progression in amyotrophic lateral sclerosis". In: *Annals of Clinical and Translational Neurology* 3.11, pp. 866–875. DOI: 10.1002/acn3.348.

Ten Have, T. R. et al. (Mar. 1998). "Mixed effects logistic regression models for longitudinal binary response data with informative drop-out". eng. In: *Biometrics* 54.1, pp. 367–383.

Thakore, N. J. et al. (Nov. 2018). "Deconstructing progression of amyotrophic lateral sclerosis in stages: a Markov modeling approach". eng. In: *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration* 19.7-8, pp. 483–494. DOI: 10.1080/21678421.2018.1484925.

Thakore, N. J. et al. (July 2019). "Variation in noninvasive ventilation use in amyotrophic lateral sclerosis". eng. In: *Neurology* 93.3, e306–e316. DOI: 10.1212/WNL.0000000000007776.

Thomadakis, C. et al. (Jan. 2024). "Joint modeling of longitudinal and competing-risk data using cumulative incidence functions for the failure submodels accounting for potential failure cause misclassification through double sampling". In: *Biostatistics* 25.1, pp. 80–97. DOI: 10.1093/biostatistics/kxac043.

Tombaugh, T. N. and N. J. McIntyre (1992). "The Mini-Mental State Examination: A Comprehensive Review". en. In: *Journal of the American Geriatrics Society* 40.9, pp. 922–935. DOI: 10.1111/j.1532-5415.1992.tb01992.x.

Tong, L., P. Liu, and E. Pena (Jan. 2022). *Joint Dynamic Models and Statistical Inference for Recurrent Competing Risks, Longitudinal Marker, and Health Status*. en.

Toth, M. J. et al. (Feb. 2000). "Effect of menopausal status on body composition and abdominal fat distribution". eng. In: *International Journal of Obesity and Related Metabolic Disorders: Journal of the International Association for the Study of Obesity* 24.2, pp. 226–231. DOI: 10.1038/sj.ijo.0801118.

Traynor, B. J. et al. (Apr. 2003). "An outcome study of riluzole in amyotrophic lateral sclerosis". In: *Journal of Neurology* 250.4, pp. 473–479. DOI: 10.1007/s00415-003-1026-z.

Trojsi, F. et al. (2012). "Clinical features and lifestyle of patients with amyotrophic lateral sclerosis in Campania: brief overview of an Italian database". eng. In: *Annali dell'Istituto Superiore Di Sanita* 48.3, pp. 287–291. DOI: 10.4415/ANN_12_03_09.

Trojsi, F. et al. (Dec. 2021). "Between-sex variability of resting state functional brain networks in amyotrophic lateral sclerosis (ALS)". eng. In: *Journal of Neural Transmission (Vienna, Austria: 1996)* 128.12, pp. 1881–1897. DOI: 10.1007/s00702-021-02413-0.

Tseng, Y.-K., F. Hsieh, and J.-L. Wang (Sept. 2005). "Joint modelling of accelerated failure time and longitudinal data". en. In: *Biometrika* 92.3, pp. 587–603. DOI: 10.1093/biomet/92.3.587.

Tukey, J. W. (1949). "Comparing Individual Means in the Analysis of Variance". In: *Biometrics* 5.2, pp. 99–114. DOI: 10.2307/3001913.

Tysnes, O. B., S. E. Vollset, and J. A. Aarli (May 1991). "Epidemiology of amyotrophic lateral sclerosis in Hordaland county, western Norway". eng. In: *Acta Neurologica Scandinavica* 83.5, pp. 280–285. DOI: 10.1111/j.1600-0404.1991.tb04701.x.

University Hospital, Lille (Nov. 2022). *Study of Predictive Factors of Progression of Motor Neurone Disease Prognosis and Endophenotype Biomarkers French Database Set up*. Clinical trial registration NCT02360891. clinicaltrials.gov.

Valliani, A. A.-A., D. Ranti, and E. K. Oermann (Aug. 2019). "Deep Learning and Neurology: A Systematic Review". In: *Neurology and Therapy* 8.2, pp. 351–365. DOI: 10.1007/s40120-019-00153-8.

Vandoorne, T., K. De Bock, and L. Van Den Bosch (Apr. 2018). "Energy metabolism in ALS: an underappreciated opportunity?" eng. In: *Acta Neuropathologica* 135.4, pp. 489–509. DOI: 10.1007/s00401-018-1835-x.

Vasta, R. et al. (Aug. 2020). "Plateaus in amyotrophic lateral sclerosis progression: results from a population-based cohort". eng. In: *European Journal of Neurology* 27.8, pp. 1397–1404. DOI: 10.1111/ene.14287.

Vehtari, A., A. Gelman, and J. Gabry (Sept. 2017). "Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC". en. In: *Statistics and Computing* 27.5, pp. 1413–1432. DOI: 10.1007/s11222-016-9696-4.

Vergonjeanne, M. et al. (Oct. 2020). "Predictive factors for gastrostomy at time of diagnosis and impact on survival in patients with amyotrophic lateral sclerosis". eng. In: *Clinical Nutrition (Edinburgh, Scotland)* 39.10, pp. 3112–3118. DOI: 10.1016/j.clnu.2020.01.018.

Vianello, A. et al. (June 2011). "Survival and quality of life after tracheostomy for acute respiratory failure in patients with amyotrophic lateral sclerosis". eng. In: *Journal of Critical Care* 26.3, 329.e7–14. DOI: 10.1016/j.jcrc.2010.06.003.

Villemagne, V. L. et al. (Apr. 2013). "Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study". eng. In: *The Lancet. Neurology* 12.4, pp. 357–367. DOI: 10.1016/S1474-4422(13)70044-9.

Virtanen, P. et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". In: *Nature Methods* 17, pp. 261–272. DOI: 10.1038/s41592-019-0686-2.

Visser, J. et al. (Apr. 2007). "Disease course and prognostic factors of progressive muscular atrophy". eng. In: *Archives of Neurology* 64.4, pp. 522–528. DOI: 10.1001/archneur.64.4.522.

Wang, J. and S. Luo (Nov. 2019). "Joint modeling of multiple repeated measures and survival data using multidimensional latent trait linear mixed model". en. In: *Statistical Methods in Medical Research* 28.10-11, pp. 3392–3403. DOI: 10.1177/0962280218802300.

Watanabe, S. (2010). "Asymptotic Equivalence of Bayes Cross Validation and Widely Applicable Information Criterion in Singular Learning Theory". en. In.

Westeneng, H.-J. et al. (May 2018). "Prognosis for patients with amyotrophic lateral sclerosis: development and validation of a personalised prediction model". eng. In: *The Lancet. Neurology* 17.5, pp. 423–433. DOI: 10.1016/S1474-4422(18)30089-9.

Westeneng, H.-J. et al. (May 2021). "Associations between lifestyle and amyotrophic lateral sclerosis stratified by C9orf72 genotype: a longitudinal, population-based, case-control study". eng. In: *The Lancet. Neurology* 20.5, pp. 373–384. DOI: 10.1016/S1474-4422(21)00042-9.

Wong, C. et al. (Oct. 2021). "Clinical trials in amyotrophic lateral sclerosis: a systematic review and perspective". In: *Brain Communications* 3.4, fcab242. DOI: 10.1093/braincomms/fcab242.

Yang, Z. et al. (Nov. 2023). *Personalized Biopsy Schedules Using an Interval-censored Cause-specific Joint Model*. DOI: 10.48550/arXiv.2209.00105.

Yeo, I. and R. A. Johnson (Dec. 2000). "A new family of power transformations to improve normality or symmetry". In: *Biometrika* 87.4, pp. 954–959. DOI: 10.1093/biomet/87.4.954.

Yiannopoulou, K. G. et al. (Dec. 2019). "Reasons for Failed Trials of Disease-Modifying Treatments for Alzheimer Disease and Their Contribution in Recent Research". In: *Biomedicines* 7.4, p. 97. DOI: 10.3390/biomedicines7040097.

Young, A. L. et al. (Jan. 2024). "Data-driven modelling of neurodegenerative disease progression: thinking outside the black box". en. In: *Nature Reviews Neuroscience*, pp. 1–20. DOI: 10.1038/s41583-023-00779-6.

Zahra, W. et al. (2020). "The Global Economic Impact of Neurodegenerative Diseases: Opportunities and Challenges". en. In: *Bioeconomy for Sustainable Development*. Ed. by C. Keswani. Singapore: Springer, pp. 333–345. ISBN: 9789811394317. DOI: 10.1007/978-981-13-9431-7_17.

Zhang, M.-J., X. Zhang, and T. H. Scheike (May 2008). "Modeling cumulative incidence function for competing risks data". In: *Expert review of clinical pharmacology* 1.3, pp. 391–400. DOI: 10.1586/17512433.1.3.391.

Zoccolella, S. et al. (Mar. 2007). "Riluzole and amyotrophic lateral sclerosis survival: a population-based study in southern Italy". en. In: *European Journal of Neurology* 14.3, pp. 262–268. DOI: 10.1111/j.1468-1331.2006.01575.x.