Mixture of random time-varying coefficients and longitudinal factor analysis models and their application to chronic pain multidimensional assessment

Amine Ounajim

THÈSE de DOCTORAT

Pour l'obtention du grade de
DOCTEUR DE L'UNIVERSITÉ DE POITIERS
Faculté des sciences fondamentales et appliquées
Laboratoire de mathématiques et applications - LMA (Poitiers) UMR 7348 CNRS
(Diplôme National - Arrêté du 25 mai 2016)

Ecole Doctorale 610
Sciences et Ingénierie des Systèmes, Mathématiques, Informatique-SISMI

Mention : Mathématiques Appliquées
Discipline : Statistique

Présentée par :
Amine Ounajim

Mixture of random time-varying coefficients and longitudinal factor analysis models and their application to chronic pain multidimensional assessment

---

Soutenue le 24/06/2022

Devant le jury composé de :

Madame Stéphanie Allassonnière Professeur, Université Paris-Descartes	**Présidente**
Monsieur François Husson Professeur, Institut Agro Rennes-Angers	**Rapporteur**
Monsieur Christophe Biernacki Professeur, Université de Lille	**Rapporteur**
Monsieur Yohann Foucher Maître de conférences des universités (HDR), Université de Nantes	**Examinateur**
Monsieur Yousri Slaoui Maitre de conférences des universités (HDR), Université de Poitiers	**Directeur de thèse**
Monsieur Pierre-Yves Louis Professeur, Université Bourgogne Franche-Comté	**Co-Directeur de thèse**
Monsieur Philippe Rigoard Professeur, Centre Hospitalier Universitaire de Poitiers et Université de Poitiers	**Co-Directeur de thèse**
Monsieur Denis Frasca Professeur, Centre Hospitalier Universitaire de Poitiers et Université de Poitiers	**Co-encadrant de thèse**

À ma mère Aicha, mon père Idder, et mon frère Zakaria.

Remerciements

Au Pr. Philippe Rigoard, je tiens à te remercier de m'avoir accueilli dans ton formidable équipe et de m'avoir guidé et montré la complexité du monde de la douleur. Grâce à toi, j'ai acquis un savoir-faire inestimable tant sur le plan professionnel que personnel. Ta rigueur, franchise, exigence et ta passion contagieuse m'ont permis de me développer et de me motiver. Je te remercie de m'avoir pris sous tes ailes il y a 6 ans.

Je tiens également à remercier mes directeurs de thèse, le Dr. Yousri Slaoui et le Pr. Pierre-Yves Louis pour leurs conseils tout au long de ce projet de thèse et pour m'avoir donné l'opportunité de rencontrer l'équipe PRISMATICS. Sans eux, rien de tout cela n'aurait été possible. Je suis très heureux et honoré d'avoir été l'un de vos étudiants.

Dr Yousri Slaoui, je ne pourrai jamais te remercier assez pour ta bienveillance et ta volonté d'aider et pour ton soutien moral et professionnel durant cette thèse. Tu as toujours été de bon conseil et de bon humeur ce qui a été un moteur important pour l'avancement de cette thèse. Je te suis très reconnaissant pour le temps conséquent que tu m'as accordé le long de mon parcours.

Pr. Pierre-Yves Louis, merci pour ta gentillesse, tes conseils précieux et la liberté que tu m'as accordée pendant la thèse. Merci aussi pour toujours trouver les mots justes pour avancer dans les moments difficiles. Ton ouverture d'esprit m'a permis d'explorer davantage d'idées. C'était toujours un plaisir et un enrichissement de discuter avec toi.

Pr Denis Frasca, étant à la fois statisticien et clinicien, c'est grâce à tes réflexions que j'ai pu trouver l'articulation entre les statistiques appliquées et le monde de la douleur. Je te remercie pour ta sympathie, de m'avoir guidée lors de la conceptualisation de cette thèse et de m'avoir permis d'avoir une vision plus claire du projet.

Pr Stéphanie Allassonnière, c'est un très grand honneur pour moi que vous ayez accepté de présider le jury de ma thèse. Votre grande expérience en intelligence artificielle appliquée à la médecine et à l'aide à la décision clinique m'apportera beaucoup de connaissance.

Je suis très reconnaissant au Pr Christophe Biernacki et au Pr François Husson pour l'intérêt que vous portez à mon travail. Je vous remercie d'avoir accepté de rapporter ce travail. C'est un grand honneur pour moi de pouvoir bénéficier de votre expertise sur les modèles de mélange et sur l'analyse exploratoire de données multidimensionnelles.

Dr. Yohann Foucher, C'est un honneur et un plaisir pour moi que vous ayez accepté d'examiner ma thèse. Les brèves discussions que nous avons eues ensemble ont été un plaisir en raison de l'intérêt de votre travail et du large éventail d'applications médicales qu'il apporte ainsi que de votre bienveillance et de votre humour.

Je tiens à exprimer ma gratitude et mes remerciements à tous ceux que j'ai rencontrés au cours de ce petit voyage.

Merci à toute l'équipe PRISMATICS de m'avoir accueilli comme une seconde famille. Merci Manuel pour ta gentillesse, ta patience et d'être toujours là pour nous guider. Merci Max pour ta rigueur et tes précieux conseils sur la rédaction et la communication scientifique ainsi que ton humour que j'apprécie. Merci à Bertille Lorgeaux et Sandrine Baron pour votre gentillesse et vos retours cliniques qui ont permis à ce travail d'avoir plus de profondeur et sans lesquels je n'aurais jamais pu accomplir ce travail. Merci à Kevin Nivole pour sa grande gentillesse et pour avoir toujours été là lorsque j'ai un problème informatique. Merci à Mathilde Many pour sa gentillesse, Lucie Lampert pour sa serviabilité, et Sarah Borel pour son enthousiasme. Je remercie également Pascaline Langlois, avec qui j'ai eu l'honneur de collaborer, pour sa gentillesse et son assiduité.

Je remercie également mes anciens collègues à PRISMATICS. Merci à Katia Gourou pour sa grande gentillesse et pour ses délicieux gâteaux. Elle est comme une tante pour moi. Merci à Céline Bouchet et Claire Brandet pour leur bonne humeur et leur bienveillance. Un grand merci à Olivier Monlezun, ancien coordinateur du laboratoire, pour son aide et son soutien durant les premières années de ce travail.

Un grand merci au Dr Nicolas Naiditch, ancien doctorant en sociologie à PRISMATICS, pour toutes les discussions passionnantes sur la douleur et les statistiques, les séances de sport, et surtout pour sa précieuse amitié.

Un immense merci à Nihel Adjali pour sa prévenance, sa gentillesse, pour sa nourriture que j'ai beaucoup appréciée pendant le Ramadan et pour les moments de discussion franche sur la culture commune de nos deux pays.

Raouf Boukenna, je sais que tu n'aimes pas les remerciements mais je tiens quand même à te dire que c'était un honneur pour moi de t'avoir comme stagiaire et c'est un plaisir de t'avoir comme ami. Je ne peux pas te remercier assez pour ton aide et ton soutien.

Clémence Bernardet, merci pour ton amitié et ta gentillesse. Merci pour la fête d'anniversaire que tu as organisée avec Nico.

Merci à tous mes anciens élèves. En particulier, je tiens à remercier Abdel Kabouri et Nassar Dakkoune pour les déjeuners marocains et les moments agréables que nous avons partagés ensemble et pour les discussions intéressantes sur nos différents projets. Je remercie également Etienne Babin pour son aide précieuse lors de son stage chez PRISMATICS et pour son humeur joviale. Merci également à Clément, désormais collègue au CHU de Poitiers, pour les discussions enrichissantes que nous avons eues ensemble.

Je tiens à remercier tous mes collègues doctorants au LMA. Je remercie Rachad Bentbib, le meilleur voisin de bureau, pour sa prévenance et sa gentillesse. Clesh Elion, un collègue bienveillant qui nous a malheureusement quitté trop tôt. Je remercie également Léo Grill, Sahar Slama, Abir Elhaj, Nesrine Chelbi, Oumaima Ben Mrad, Angélique Perrillat, Paolo Menegatti, Clément Chesseboeuf, et tous les autres doctorants que j'ai eu l'honneur de rencontrer au LMA. Je remercie également mes anciens professeurs du LMA Hermine Biermé, Farida Enikeeva et Julien Michel.

Un grand merci aux internes et aux médecins avec lesquels j'ai eu l'honneur de collaborer. Je remercie particulièrement le Dr Adrien Delaveau, le Dr Abdollah-Yassine Moufid et le Dr Tanguy Vandeuvre pour m'avoir beaucoup appris sur la chirurgie orthopédique.

Et pour garder le meilleur pour la fin. Je tiens à remercier toute ma famille, en particulier ma mère, qui a toujours été là malgré la distance. Elle m'a toujours soutenu et encouragé. Je remercie mon père, pour m'avoir toujours soutenu dans les moments difficiles. Je ne remercierai jamais assez mes parents. Enfin, je remercie mon grand frère, Zakaria, sans qui je ne serais pas là aujourd'hui. Il m'a transmis son intérêt pour les mathématiques et a toujours été là pour moi. Merci frérot.

Contribution scientifique

Publications

Dans le cadre de la thèse

1. Ounajim A, Billot M, Louis P-Y, Slaoui Y, Frasca D, Goudman L, Roulaud M, Naiditch N, Lorgeoux B, Baron S, Nivole K, Many M, Adjali N, Page P, Bouche B, Charrier E, Poupin L, Rannou D, de Montgazon GB, Roy-Moreau B, Wood C, Rigoard R, David R, Moens M, Rigoard P. Finite Mixture Models Based on Pain Intensity, Functional Disability and Psychological Distress Composite Assessment Allow Identification of Two Distinct Classes of Persistent Spinal Pain Syndrome after Surgery Patients Related to Their Quality of Life. Journal of Clinical Medicine. 2021 ; 10(20) :4676. https ://doi.org/10.3390/Ounajim2021
2. Ounajim A, Slaoui Y, Louis P-Y, Billot M, Frasca D, Rigoard P. Mixture of varying-coefficient models with random effects processes for intra and inter-subject effect estimation (Soumis en révision).
3. Ounajim A, Slaoui Y, Louis P-Y, Billot M, Frasca D, Rigoard P. Mixture of longitudinal factor analysis with time-variant loadings for heterogeneous longitudinal multivariate data (En préparation).

Autres publications

1. Ounajim A, Billot M, Goudman L, Louis P, Slaoui Y, *et al.* Machine Learning Algorithms Provide Greater Prediction of Response to SCS Than Lead Screening Trial : A Predictive AI-Based Multicenter Study. *J. Clin. Med.* 2021, 10, 4764. doi.org/10.3390/jcm10204764
2. Langlois P, Perrochon A, David R, *et al.*. Hypnosis to manage musculoskeletal and neuropathic chronic pain : a systematic review and meta-analysis. Neuroscience and Biobehavioral Reviews (accepted). 2022.
3. Rigoard P, Ounajim A, Goudman L, *et al.* The Added Value of Subcutaneous Peripheral Nerve Field Stimulation Combined with SCS, as Salvage Therapy, for Refractory Low Back Pain Component in Persistent Spinal Pain Syndrome Implanted Patients : A Randomized Controlled Study (CUMPNS Study) Based on 3D-Mapping Composite Pain Assessment. Journal of Clinical Medicine. 2021 ; 10(21) :5094. doi.org/10.3390/jcm10215094.

4. Rigoard P, Ounajim A, Goudman L, et al. The Challenge of Converting "Failed Spinal Cord Stimulation Syndrome" Back to Clinical Success, Using SCS Reprogramming as Salvage Therapy, through Neurostimulation Adapters Combined with 3D-Computerized Pain Mapping Assessment : A Real Life Retrospective Study. J Clin Med. 2022 ;11(1) :272. doi :10.3390/jcm11010272.

5. Rigoard P, Ounajim A, Goudman L, Louis P, Slaoui Y, *et al.* A Novel Multi-dimensional Clinical Response Index Dedicated to Improving Global Assessment of Pain in Patients with Persistent Spinal Pain Syndrome After Spinal Surgery, Based on a Real-life Prospective Multicentric Study (PREDIBACK) and Machine Learning. Journal of Clinical Medicine. 2021 ; 10(21) :4910. doi.org/10.3390/jcm10214910.

6. Naiditch N, Billot M, Moens M, et al. Persistent Spinal Pain Syndrome Type 2 (PSPS-T2), a Social Pain ? Advocacy for a Social Gradient of Health Approach to Chronic Pain. J Clin Med. 2021 ;10(13) :2817. doi :10.3390/jcm10132817

7. Rigoard P, Billot M, Ingrand P, et al. How Should we Use Multicolumn Spinal Cord Stimulation to Optimize Back Pain Spatial Neural Targeting ? A Prospective, Multicenter, Randomized, Double-Blind, Controlled Trial (ESTIMET Study). Neuromodulation. 2021 ;24(1) :86-101. doi :10.1111/ner.13251.

8. Naiditch N, Billot M, Goudman L, Cornet P, *et al.* Professional status of Persistent Spinal Pain Syndrome patients after spinal surgery (PSPS-T2) : What really matters ? A prospective study introducing the concept of “Adapted Professional Activity” inferred from clinical, psychological and social influence. Journal of Clinical Medicine. 2021 ; 10(21) :5055. doi.org/10.3390/jcm10215055.

9. Rigoard P, Roulaud M, Goudman L, *et al.* Comparison of Spinal Cord Stimulation vs. Dorsal Root Ganglion Stimulation vs. Association of Both in Patients with Refractory Chronic Back and/or Lower Limb Neuropathic Pain : An International, Prospective, Randomized, Double-Blinded, Crossover Trial (BOOST-DRG Study). Medicina (Kaunas). 2021 ;58(1) :7. doi :10.3390/medicina58010007.

10. Billot M, Naiditch N, Brandet C, et al. Comparison of conventional, burst and high-frequency spinal cord stimulation on pain relief in refractory failed back surgery syndrome patients : study protocol for a prospective randomized double-blinded cross-over trial (MULTIWAVE study). Trials. 2020 ;21(1) :696. Published 2020 Aug 3. doi :10.1186/s13063-020-04587-6

11. Vendeuvre T, Brèque C, Severyns M, Ounajim A, Richer JP, Rigoard P, Germaneau A. Interest of PMMA filing to primary stabilization of the minimaly invasive osteosynthesis of schatzker type II tibial Plateau fractures, Computer Methods in Biomechanics and Biomedical Engineering. 2019 ; 22 :sup1, S289-S290, DOI : 10.1080/10255842.2020.1714916

12. Moufid AY, Cloche T, Ghailane S, Ounajim A, Vendeuvre T, Gille O. Mismatch between rod bending and actual post-operative lordosis in lumbar arthrodesis with poly axial screws. Orthop Traumatol Surg Res. 2019 ;105(6) :1143-1148. doi :10.1016/j.otsr.2019.03.003

13. Delaveau A, Saint-Genez F, Gayet LE, Paccalin M, Ounajim A, Vendeuvre T. Impact of time to surgery in upper femoral fracture in orthogeriatrics. Orthop Traumatol Surg Res. 2019 ;105(5) :975-978. doi :10.1016/j.otsr.2019.04.018
14. Rigoard P, Ounajim A, North RB. Questioning Prediction of Lumbar Spine Surgery Outcome-Why We Need to Pay Attention. JAMA Surg. 2018 ;153(11) :1061-1062. doi :10.1001/jamasurg.2018.2500

Communications orales

1. Ounajim A, Slaoui Y, Louis P-Y, Billot M, Frasca D et Rigoard P. Mélange de modèle à coefficients variables incluant des effets aléatoires. *Algorithmes Stochastiques, Modélisation Statistiques et Applications : Quatrième Rencontre Poitiers-Bordeaux*, En ligne, 11 décembre 2020.
2. Ounajim A, Slaoui Y, Louis P-Y, Billot M, Frasca D et Rigoard P. Mixture of varying coefficients models with random effects. 32^{th} *European meeting of Statisticians (EMS 2019)*, Palermo Italie, 25 juillet 2019.
3. Ounajim A. Roulaud M, Naiditch N, Rigoard P. Présentation et résultats préliminaires de l'étude PREDIBACK – Etude prospective multicentrique sur l'optimisation de la prise en charge des patients souffrant de lombo-radiculalgies post-opératoires. *Journée BOLIPO, Douleur : échange entre professionnels*, La Rochelle, 22 Mars 2019.
4. Ounajim A. Application des méthode de Machine Learning pour prédire l'éfficacité de la neurostimulation chez les patients avec des douleurs chroniques post-opératoires. *Statistique appliquée à la biologie et à la santé*, Rencontre Poitiers, 1 février 2018.

Table des matières

Introduction générale

Contexte statistique

Les développements récents et les avancées technologiques dans les méthodes de collecte de données ont permis aux chercheurs de réaliser de grandes études de cohorte avec un suivi longitudinal à long terme [1]. Par conséquent, les méthodes statistiques appliquées à ces données présentent un grand intérêt et ont gagné en popularité au cours des dernières années. Cela a même conduit à l'attribution du prix international de statistique en 2021 au professeur Nan Laird, biostatisticienne de l'Université de Harvard, pour ses travaux sur les méthodes d'analyse des données longitudinales complexes.
Les données longitudinales multivariées jouent un rôle essentiel dans divers domaines de recherche, notamment les sciences médicales, sociales et comportementales. Elles permettent en effet aux chercheurs de tester de multiples hypothèses, comme l'identification de la variation dans le temps d'une variable clinique ou de la prévalence d'une maladie, ou encore d'estimer l'évolution dans le temps de l'effet d'une thérapie sur un ou plusieurs critères d'évaluation [2]. La robustesse des inférences obtenues à partir de l'analyse de données longitudinales induit une complexification croissante des méthodes de traitement statistique [3]. Cette complexité découle de la structure des données longitudinales et de leurs propriétés statistiques incluant une dépendance temporelle intra-individuelle. Le design expérimental par lequel les données sont collectées (*e.g.* le temps entre les visites, l'absence de données, le nombre d'observations par sujet, ...) et la structure de variance-covariance sont quelques-unes des propriétés qui doivent être prises en compte afin de mener une analyse appropriée et précise des données longitudinales.
Plusieurs méthodes d'analyse des données longitudinales multivariées sont disponibles dans la littérature, notamment les modèles mixtes multivariés [4, 5] et l'analyse factorielle pour les données longitudinales (*longitudinal factor analysis*). Il existe également les modèles linéaires mixtes à facteurs latents [6, 7], les modèles de courbes de facteurs (*curve of factors model*) et les modèles de facteur de courbes (*factor of curves model*) [8], l'analyse factorielle dynamique (*dynamic factor analysis*) [9], le modèle d'équations structurelles dynamiques (*dynamic structural equations modeling*) et enfin, les modèles à processus latents [10]. Ces méthodes acceptent généralement plusieurs variables observées comme variables de sortie. Ces variables ne sont généralement qu'une représentation bruitée de facteurs latents sous-jacents qui ne peuvent être observés ou mesurés. Par exemple, la dépression est « mesurée » à l'aide d'un ensemble d'items (variables observés) contenant une erreur de mesure. Une proportion de l'information contenue dans ces items est

représentée par le facteur latent, qui est dans cet exemple, la sévérité de l'état dépressif. En général, pour étudier l'évolution de plusieurs indicateurs bruités chez des patients échantillonnés dans une population homogène, il est important de se concentrer sur les tendances longitudinales parmi les variables latentes en utilisant une modélisation conjointe fondée sur les structures de la variance-covariance plutôt que de considérer chaque indicateur séparément. Cette approche permet de réduire le nombre d'analyses à effectuer et donc d'aboutir à des conclusions plus généralisables et robustes.

Plusieurs des méthodes présentées précédemment partagent le même modèle de mesure (measurement model) qui est l'équation représentant la relation entre les variables mesurées/observées et les facteurs latents. En revanche, le modèle structurel (structural model) représentant la relation entre les facteurs latents eux-mêmes et leur évolution dans le temps, diffère selon les modèles. Par exemple, dans le modèle factoriel dynamique, un modèle autorégressif est utilisé pour modéliser la relation entre les structures factorielles précédentes et actuelles, alors que dans le modèle linéaire mixte à facteurs latents, les facteurs latents sont modélisés comme le résultat d'un modèle à effets mixtes dans lequel des fonctions polynomiales du temps peuvent être considérées pour représenter l'évolution longitudinale des facteurs latents. Ces modèles diffèrent également dans la prise en compte d'une structure factorielle variante ou invariante (saturations factorielles dépendantes du temps). L'invariance des facteurs est une hypothèse classique dans l'analyse factorielle puisque la structure des construits latents est supposée être invariante afin de permettre la comparabilité des facteurs dans le temps. Toutefois, dans certains domaines applicatifs, il est connu que la structure des construits latents (des variables qui ne peuvent pas être mesurées) change, ce qui peut être dû, par exemple, à l'évolution culturelle [11] ou à celle du stade de la maladie.

Une autre hypothèse plus restrictive des modèles discutés précédemment et des modèles d'analyse factorielle en général, est l'invariance de la structure factorielle parmi des sous-groupes de sujets. Cette hypothèse suppose que la population étudiée est homogène en ce qui concerne la structure des variables mesurées. En sciences humaines et de la santé, il est possible que la structure des facteurs varie entre les sous-populations. L'impact des variables mesurées sur les construits latents sous-jacents peut varier en fonction d'autres caractéristiques latentes et observées (*e.g.* sexe, âge, niveau d'études, diagnostic, comportement). Pour résoudre ce problème, certains auteurs ont proposé d'utiliser un mélange de modèles d'analyses factorielles [12]. Ces modèles estiment différentes saturations factorielles pour différentes sous-populations, qui sont représentées par des classes latentes. À notre connaissance, aucune extension aux données longitudinales de ces modèles n'a été proposée.

Dans cette thèse nous proposons deux nouveaux types de modèles d'analyses de données longitudinales. Le premier est un mélange de modèles de régression à coefficients variables (non-paramétriques) avec des effets aléatoires représentés par des processus stochastiques Gaussiens. Ce modèle permet d'estimer l'effet de plusieurs variables explicatives longitudinales sur une variable cible longitudinal continue, en utilisant des données où les observations sont tirées de plusieurs populations hétérogènes inconnues. Ce modèle nous permet de modéliser l'effet évolutif des caractéristiques des patients et des traitements sur leur résultat. Le deuxième type de modèles proposé dans cette thèse est le mélange de modèles d'analyses factorielles longitudinales. Ce modèle est

une extension du modèle mixte linéaire à facteurs latents proposé par An et al. [7]. Il permet de supposer une variation dans le temps de la structure des facteurs latents (structure représentée par les saturations factorielles). Ce modèle permet aussi, grâce à l'inclusion d'un mélange, de supposer qu'il existe plusieurs structures factorielles différentes associées à des sous-populations inconnues (classes latentes). Enfin, il permet de décrire plusieurs variables longitudinales cibles en fonction de plusieurs variables longitudinales explicatives grâce une analyse factorielle dont la structure varie dans le temps et entre différents groupes de sujets.

Contexte clinique

Douleurs chronique, causes et conséquence

Les douleurs chroniques (douleurs qui persistent depuis plus de 3 mois) sont complexes, épuisantes et représentent un fardeau important pour le patient et pour la société [13]. Ces douleurs apparaissent généralement suite à une maladie ou un traumatisme. Les douleurs chroniques sont une symptomatologie qui regroupe plusieurs étiologies qui disposent de leur propre taxonomie et de leur définition médicale [14]. Cette hétérogénéité de la population des douloureux chroniques rend difficile son évaluation et sa prise en charge.
Malgré le fardeau que les douleurs chroniques ont sur les patients et la société, 70% des patients douloureux ne sont pas traités de manière adéquate d'après le livre blanc de la douleur 2017 établi par la société française d'étude et de traitement de la douleur [15]. La prise en charge de référence pour les douleurs chroniques est l'utilisation des médicaments antalgiques incluant les opioïdes [16]. En France, selon l'observatoire français des médicaments antalgiques, l'usage des opioïdes conduit à une moyenne de quatre décès (par 1000000 individus) par semaine [17]. **Le taux d'overdose est plus important chez les douloureux chroniques que chez les usagers de drogues illégales.** Les traitements inadaptés ont souvent des effets secondaires difficiles à vivre, notamment des dépendances graves qui ont un impact désastreux sur les activités quotidiennes et l'état psychologique des patients [18].
Selon les auteurs d'une revue de la littérature publiée dans le journal *the New England Journal of Medicine* [19], l'idée selon laquelle les antalgiques opioïdes sont efficaces et sûrs reposerait sur l'efficacité perçue de ces médicaments pour traiter les douleurs aiguës (douleurs inférieures à 3 mois) et les douleurs en soins palliatifs. Cette efficacité est évaluée en se basant sur le principe que la dose correcte d'un opioïde est celle qui permet de réduire l'intensité de la douleur, telle que mesurée par une échelle d'intensité de la douleur comprise entre 0 (aucune douleur) et 10 (la pire douleur imaginable). Ceci soulève deux problématiques liées à l'évaluation des douleurs chroniques et à leur prise en charge. La première problématique est de considérer une réduction de l'intensité de la douleur comme la finalité absolue d'une prise en charge des douleurs. La deuxième problématique est de ne pas prendre en compte l'évolution temporelle des douleurs. La littérature montre que la structure liant la nociception de la douleur et les émotions devient plus forte chez les douloureux chroniques que chez les douloureux aiguës [20–22]. Par conséquent, juger qu'un traitement est efficace pour les douleurs chroniques en se fondant sur son efficacité pour les douleurs aiguës est inadéquat. En outre, chez les patients douloureux chroniques, la

souffrance associée aux douleurs génère une détérioration de la capacité fonctionnelle, de la stabilité sociale ainsi que de la qualité de vie. Or ces dimensions ne sont pas forcément améliorées par la réduction de l'intensité de la douleur [23].
Nous allons par la suite présenter deux arguments majeurs qui expliquent l'inadéquation de l'évaluation actuelle des douleurs chroniques:

Argument 1 : La multidimensionnalité de l'évaluation des douleurs chroniques

Les douleurs chroniques ne sont pas juste une perception sensorielle associée à des potentielles lésions des tissues, mais aussi une expérience associée à une détresse émotionnelle et un handicap physique qui conduisent à une détérioration de la qualité de vie [24]. Par conséquent, la réduction de l'intensité de la douleur n'est pas forcément le bon/le seul objectif à atteindre pour le traitement des douleurs chroniques. Les opioïdes sont utilisés de plus en plus fréquemment et à des doses croissantes pour tenter de réduire les scores de l'intensité de la douleur - tout en augmentant les taux d'effets toxiques des médicaments, en exposant des populations vulnérables à des risques (la dépendance, une surdose, ...) et en ne parvenant pas toujours à améliorer les conséquences des douleurs chroniques sur la vie quotidienne [25]. La poursuite d'une diminution de l'intensité de la douleur au prix d'une détérioration de la capacité fonctionnelle et de la qualité de vie n'est surement pas l'approche optimale à poursuivre.

La complexité et l'hétérogénéité de la population rend difficile l'identification de traitements efficaces et peut conduire, comme dans le cas des opioïdes, à une prise en charge complétement déconnectée des besoins réels des patients comme se mouvoir, se vêtir, se laver et se socialiser. La douleur n'est en général encore évaluée que par son intensité, à l'aide d'une échelle numérique allant de 0 à 10, ce qui ne permet pas de prendre en compte les différentes dimensions des douleurs telles que la capacité fonctionnelle et la détresse psychologique. Paradoxalement, certains patients déclarent avoir une douleur très intense, qui reste pourtant acceptable à leurs yeux [26]. Au niveau cortical, il a été démontré que la chronicisation de la douleur (passage des douleurs aiguës aux douleurs chroniques) génère des altérations de la connectivité entre les zones associées à la sensation de douleur, aux émotions et à la cognition [27]. Malgré la reconnaissance par la littérature de l'aspect multidimensionnel des douleurs [28], son évaluation en pratique clinique reste à ce jour partielle [29–31]. Concernant la recherche clinique sur les douleurs chroniques, des mesures qui incorporent plusieurs dimensions des douleurs chroniques ont été développées [28]. Toutefois, 60% des études utilisent l'échelle de l'intensité de la douleur comme le critère principal et 20% des études utilisent uniquement l'intensité de la douleur comme outil d'évaluation [32, 33]. Les essais cliniques restent la source la plus importante de preuves sur l'efficacité des traitements. Cependant, comme certains traitements sont conçus pour améliorer la qualité de vie et la capacité fonctionnelle, il semble que nous réduisons l'impact des recherches en limitant l'évaluation des douleurs principalement à leur intensité. Pourtant, cela semble être l'état actuel des choses dans la majorité des essais cliniques. Une des raisons qui explique que la communauté médicale continue d'utiliser l'échelle d'évaluation de la douleur est sa simplicité d'utilisation. Le développement d'un outil multidimensionnel permettant de combiner les différentes dimensions des douleurs (sensorielle, fonctionnelle, psychologique, ...) d'une manière simple et rapide, pourrait améliorer l'évaluation des douleurs et les recommandations thérapeutiques.

Argument 2 : Caractère longitudinal de la douleur

La douleur aiguë est la prise de conscience d'un signal nocif provenant d'un tissu récemment endommagé. Son intensité varie en fonction des processus inflammatoires, de la cicatrisation des tissus et du mouvement. La durée d'apparition des douleurs aiguës est l'une de ses principales caractéristiques (durée inférieure à 3 mois). Les douleurs aiguës font partie de la vie quotidienne et sont généralement peu préoccupantes car chacun sait par expérience comment en interpréter la cause et en prévoir l'issue. Cette douleur peut être atroce, mais elle est mieux tolérée et cause moins de souffrance parce qu'elle a une durée limitée et qu'elle peut être nécessaire pour atteindre un objectif valable, comme une bonne performance sportive ou un accouchement. En revanche, dans le cas des douleurs chroniques, le sentiment persistant d'impuissance et de souffrance ainsi que l'impact à long terme sur la vie quotidienne génèrent de la souffrance chez les patients [34].
Actuellement, la littérature reconnait que les douleurs chroniques différent des douleurs aiguës par rapport à ses causes, sa psychopathologie, sa portée sociale et son impact sur l'activité cérébrale [35, 36]. Il a également été démontré que les corrélations entre les différentes dimensions des douleurs deviennent également plus fortes avec le temps, non seulement en passant des douleurs aiguës aux douleurs chroniques, mais aussi au cours de l'évolution des douleurs chroniques [22]. Ceci peut être dû à l'accumulation à long terme des impacts bio psycho sociaux générés par les douleurs chroniques et aux altérations des mécanismes cérébraux liant la douleur à l'émotion. Par conséquent, comme nous l'avons discuté précédemment, l'emprunt des principes d'évaluation et de prise en charge des douleurs aiguës et des douleurs en soins palliatifs pour traiter les douleurs chroniques n'est pas adéquat. L'évaluation des douleurs chroniques doit prendre en compte le caractère évolutif/variant de la structure comprenant les différentes dimensions de la douleur.

Objectifs de la thèse

Dans cette thèse, nous nous focaliserons sur **les douleurs chroniques post-chirurgie du rachis**. Les douleurs chroniques post-chirurgie du rachis (Chronic Pain After Spinal Surgery ou CPASS), récemment incluses dans la $11^{ème}$ révision de la Classification Internationale des Maladies (CIM-11), est définie comme des douleurs chroniques apparaissant après une intervention chirurgicale et persistant au-delà du processus de guérison [37]. Ce type de douleurs chroniques affecte environ **20% des patients subissant une chirurgie du rachis lombaire.** La prise en charge de ces patients représente un problème majeur de santé publique [38]. En tant que sous-classification des douleurs chroniques, les douleurs chroniques après une chirurgie du rachis représentent également les caractéristiques discutées ci-dessus : la multidimensionnalité de son évaluation et le caractère longitudinal de l'association entre ses dimensions. Ces deux caractéristiques ne sont pas actuellement prises en compte par les outils d'évaluation des douleurs. Afin d'y remédier, **nous développerons de nouvelles techniques d'analyses statistiques multidimensionnelles longitudinales** pour atteindre l'objectif suivant : **Développer un modèle statistique permettant d'obtenir des scores multidimensionnels d'évaluation de la douleur dont les composantes évoluent avec le temps et selon les différentes sous-populations hétérogènes.** Ce type de modèle nous permettra, à un instant donné, de générer un score pour chaque patient en fonction d'une combinaison pondérée de plusieurs critères. Puisque les mécanismes et

l'impact des douleurs évoluent dans le temps, les pondérations de cette combinaison varient dans le temps. En outre, différentes combinaisons sont utilisées pour différentes sous-populations et pour différents instants du parcours du patient.

Déroulement de la thèse

Nous commencerons dans le chapitre 2 par introduire les différentes notions et méthodes statistiques proposées dans la littérature pour traiter les données longitudinales hétérogènes. Les méthodes discutées incluent les modèles à effets mixtes [3], les modèles à coefficient variables [39] et les modèles d'analyses factorielles longitudinales [6, 7]. Nous introduirons aussi la notion de mélange de modèles qui sont des méthodes qui permettent de prendre en compte l'hétérogénéité des données [40].
Le chapitre 3 consistera en une application d'un mélange fini de modèles à effets mixtes sur une base de données de patients avec des douleurs chroniques après une chirurgie du rachis [41]. Ce chapitre illustrera l'utilité des modèles de mélange dans le cadre d'analyse de données longitudinales. Notre objectif dans ce chapitre sera d'extraire des clusters de patients basés sur l'impact de l'incapacité fonctionnelle, de l'intensité de la douleur et de la détresse psychologique sur leur qualité de vie et d'identifier les caractéristiques sociodémographiques, comportementales et cliniques associées à ces clusters. L'analyse dans cette étude a été réalisée à partir d'un ensemble de données de 200 patients de l'étude observationnelle multicentrique française PREDIBACK incluant des patients souffrant de douleurs chroniques après une chirurgie du rachis avec un suivi d'un an. L'identification de ces classes permettra une meilleure compréhension de la façon dont la qualité de vie des patients sera impactée par les différentes dimensions des douleurs chroniques. De plus, cela donnera une ouverture vers une évaluation personnalisée de la douleur et de la qualité de vie. Dans cette thèse, cette application nous permettra d'introduire les mélanges de modèles à effets mixtes ainsi que de montrer l'hétérogénéité de la population des patients douloureux chroniques par rapport à la manière dont ils sont impactés par la composante fonctionnelle, sensorielle et psychologique des douleurs chroniques.
Le chapitre 4 aura pour but de présenter le développement statistique d'un mélange (qui dépend du temps) de modèles à coefficients variables incluant des effets aléatoires représentés par des processus gaussiens. Ce développement découle d'une nécessité applicative due au fait que les modèles à coefficients variables ne prennent pas en compte l'hétérogénéité des données ou la corrélation intra-sujet. Ceci correspond à un biais d'estimation majeur, qui constitue une forte limitation des modèles à coefficients variables classiques dans le cadre de données longitudinales. Nous présenterons une procédure spécifique de *backfitting* pour ajuster notre modèle aux données. Une méthode de validation croisée pour la sélection des paramètres de lissage sera également présentée. Le modèle proposé sera évalué sur des données simulées et réelles, montrant que notre algorithme rassemble les sujets en groupes homogènes et permet d'obtenir de meilleures estimations des effets intra et inter-sujets. Ce modèle peut être considéré comme une version non-paramétrique du mélange de modèles à effets mixtes utilisé dans le chapitre 2. Les méthodes proposées ont été programmées sur le logiciel R (R Foundation for Statistical Computing, Vienna, Austria). Nous appliquerons ce modèle sur une base de données afin d'illustrer son utilité. Le modèle sera appliqué sur la base de données PREDIBACK afin de

valider les clusters obtenus dans le chapitre 3 et évaluer leur évolution dans le temps.
Dans le chapitre 5, nous proposerons une nouvelle méthode qui repose sur un modèle de mélange d'analyse factorielle longitudinale. Dans un cadre général, ce modèle permettra de décrire plusieurs variables longitudinales dépendantes en fonction de plusieurs variables longitudinales indépendantes grâce à une analyse factorielle dont la structure varie dans le temps. Le modèle inclut également un mélange afin de permettre aux structures factorielles de varier entre les individus. Nous appliquons le modèle à la base de données PREDIBACK où les patients sont évalués par plusieurs questionnaires mesurant plusieurs aspects comme l'intensité de la douleur, la qualité de vie, la capacité fonctionnelle, la dépression, l'anxiété, la catastrophisation et autres. Notre objectif est de réduire toutes ces dimensions en un nombre beaucoup plus restreint de facteurs latents qui représentent des concepts plus globaux, en tenant compte du caractère évolutif de la douleur et des différences inter-individuelles dans son impact.
Dans le chapitre 6, nous discuterons les travaux de cette thèse, nous présenterons les travaux de recherche futurs, puis nous terminerons avec une conclusion générale de la thèse.
Le chapitre 3 fait l'objet d'un papier accépté dans *Journal of Clinical Medicine*. Le chapitre 4 fait l'objet d'un article soumis pour publication. Le chapitre 5 fait l'objet d'un travail en cours qui sera soumis pour publication. Dans le cadre de mon activité en tant que statisticien au CHU de Poitiers, j'ai eu l'opportunité de co-rédiger 8 articles cliniques publiés (dont 1 en premier auteur disponible en annexe) et 4 soumis pour publication.

Chapitre 1

General introduction

1.1 Statistical context

Recent developments and technological advances in data collection methods have allowed researchers to conduct large cohort studies with long-term follow-up [1]. As a result, statistical methods applied to these data are of great interest and have gained popularity in recent years. This has even led to the awarding of the International Prize in Statistics in 2021 to Professor Nan Laird, a biostatistician at Harvard University, for her work on methods for analyzing complex longitudinal data.
Multivariate longitudinal data play an essential role in various fields of research, including the medical, social and behavioral sciences. They allow researchers to test multiple hypotheses, such as identifying the variation over time of a clinical variable or disease prevalence, or to estimate the evolution over time of the effect of a therapy on one or more endpoints [2]. The robustness of the inferences obtained from the analysis of longitudinal data leads to an increasing complexity of the statistical processing methods [3]. This complexity stems from the structure of the longitudinal data and their statistical properties. The experimental design by which the data are collected (*e.g.* time between visits, absence of data, number of observations per subject, ...) and the variance-covariance structure are some of the properties that must be taken into account in order to conduct an appropriate and accurate analysis of longitudinal data.
Several methods for analyzing multivariate longitudinal data are available in the literature, including multivariate mixed models [4, 5] and factor analysis for longitudinal data (longitudinal factor analysis). There are also linear mixed models with latent factors [6, 7], curve of factors model and factor of curves model [8], dynamic factor analysis [9], dynamic structural equations modeling and finally, latent process models [10]. These methods generally accept several observed variables as output variables. These variables are usually only a noisy representation of underlying latent factors that cannot be observed or measured. For example, depression is "measured" using a set of items (observed variables) containing measurement error. A proportion of the information contained in these items is represented by the latent factor, which in this example is the severity of the depressive state. In general, to study the evolution of several noisy indicators in patients sampled

from a homogeneous population, it is important to focus on longitudinal trends among the latent variables using joint modeling based on variance-covariance structures rather than considering each indicator separately. This approach reduces the number of analyses that need to be performed and thus leads to more generalizable and robust conclusions.
Several of the methods presented above share the same measurement model, which is the equation representing the relationship between the measured/observed variables and the latent factors. On the other hand, the structural model, which represents the relationship between the latent factors themselves and their evolution over time, differs according to the models. For example, in the dynamic factor model, an autoregressive model is used to model the relationship between the previous and current factor structures, whereas in the linear mixed latent factor model, the latent factors are modeled as the result of a mixed-effects model in which polynomial time functions are considered to represent the longitudinal evolution of the latent factors. These models also differ in the consideration of a variant or invariant factor structure (time-dependent factor loadings). Factor invariance is a classic assumption in factor analysis since the structure of the latent constructs is assumed to be invariant in order to allow comparability of factors over time. However, in some domains, it is well known that the structure of latent constructs changes, which may be due, for example, to cultural evolution [11] or disease stage changes.
Another more restrictive assumption of the models discussed above, and of factor analysis models in general, is the invariance of the factor structure among subgroups of subjects. This assumption assumes that the population studied is homogeneous with respect to the structure of the constructs measured. In education and health sciences, it is possible that the structure of the factors varies between subpopulations. The impact of measured variables on the underlying latent constructs may vary according to other latent and observed characteristics (e.g., gender, age, education, diagnosis, behavior). To address this problem, some authors have proposed to use a mixture of factor analysis models [12]. These models estimate different factor loadings for different subpopulations, which are represented by latent classes. To our knowledge, no extension of these models to longitudinal data has been proposed.
In this thesis, we propose two new types of models for longitudinal data analysis. The first one is a mixture of variable coefficients (non-parametric) regression models with random effects represented by stochastic Gaussian processes. This model allows us to estimate the effect of several longitudinal explanatory variables on a continuous longitudinal target variable, using data where the observations are drawn from several unknown heterogeneous populations. This model allows us to model the evolutionary effect of patient characteristics and treatments on their outcome. The second type of model proposed in this thesis is the mixed model of longitudinal factor analysis. This model is an extension of the mixed linear latent factor model proposed by An et al. [7]. It allows to assume a variation in time of the structure of the latent factors (structure represented by the factor loadings). This model also allows, thanks to the inclusion of a mixture, to assume that there are several different factorial structures associated with unknown sub-populations (latent classes). Finally, it allows the description of several target longitudinal variables as a function of several explanatory longitudinal variables through a factor analysis whose structure varies over time and between different groups of subjects.

1.2 Clinical context

1.2.1 Chronic pain, causes and consequences

Chronic pain (pain that has persisted for more than 3 months) is complex, exhausting, and a significant burden to society and especially to the patient [13]. This pain usually occurs as a result of trauma, illness, or injury. Chronic pain is a symptomatology that groups together several etiologies, each of which has its own taxonomy and medical definition [14]. This heterogeneity of the chronic pain population makes it difficult to assess and manage.
Despite the burden that chronic pain has on society and patients, 70% of pain patients are not adequately treated according to the 2017 Pain White Paper compiled by the French Society for the Study and Treatment of Pain. The reference management for chronic pain is the use of analgesic drugs including opioids [16]. In France, according to the French Observatory of Analgesic Drugs, opioid use leads to an average of four deaths (per 1000000 individual) per week [17]. The rate of overdose is greater among chronic pain sufferers than among illegal drug users. Inappropriate treatments often have difficult side effects, including severe addictions that have a disastrous impact on patients' daily activities and psychological state [18].
According to the authors of a review of the literature published in the New England Journal of Medicine [19], the idea that opioid analgesics are effective and safe is based on the perceived efficacy of these drugs in treating acute pain (pain less than 3 months) and pain in palliative care (end-of-life phase). This efficacy is assessed on the basis that the correct dose of an opioid is the one that reduces the intensity of pain, as measured by a pain intensity scale ranging from 0 (no pain) to 10 (the worst pain imaginable). This raises two issues related to the assessment and management of chronic pain. The first problem is to consider a reduction in pain intensity as the absolute goal of pain management. The second problem is not to take into account the temporal evolution of pain. The literature shows that the structure linking pain nociception and emotions becomes stronger in chronic pain patients than in acute pain patients [20–22]. Therefore, judging that a treatment is effective for chronic pain based on its effectiveness for acute pain is not adequate. Moreover, in chronic pain patients, the suffering associated with pain leads to a deterioration in functional capacity, social stability and quality of life. These dimensions are not necessarily improved by reducing the intensity of the pain [23].

1.2.2 Multidimensionality of chronic pain evaluation

Chronic pain is not just a sensory perception associated with potential tissue damage, but also an experience associated with emotional distress and physical disability that leads to a deterioration in quality of life [24]. Therefore, reducing pain intensity is not necessarily the right/only goal for the treatment of chronic pain. Opioids were being used with increasing frequency and at increasing doses in an attempt to reduce pain intensity scores - while increasing rates of drug toxicity, exposing vulnerable populations to risks (addiction, overdose, etc.), and not always improving the impact of chronic pain on daily life [25]. The pursuit of a decrease in pain intensity at the cost of a deterioration in functional capacity and quality of life is surely not the optimal approach to pursue.

The complexity and heterogeneity of the population makes it difficult to identify effective treatments and can lead, as in the case of opioids, to a management completely disconnected from the real needs of patients such as moving, dressing, washing and socializing. Pain is generally still assessed only by its intensity, using a numerical scale ranging from 0 to 10, which does not allow the different dimensions of pain such as functional capacity and psychological distress to be taken into account. Paradoxically, some patients report very intense pain, which is still acceptable to them [26]. At the cortical level, it has been demonstrated that the chronicisation of pain (passage from acute to chronic pain) generates alterations in the connectivity between the areas associated with the sensation of pain, emotions and cognition [27]. Despite the recognition by the literature of the multidimensional aspect of pain [28], its evaluation in clinical practice remains to date partial [29–31]. Regarding clinical research on chronic pain, measures that incorporate several dimensions of chronic pain have been developed. However, 60% of studies use the pain intensity scale as the primary endpoint and 20% of studies use pain intensity alone as an assessment tool [32, 33]. Clinical trials remain the most important source of evidence on the effectiveness of treatments, but because some treatments are designed to improve quality of life and functional capacity, it seems that we are reducing the impact of research by limiting the assessment of pain primarily to its intensity. Yet this seems to be the current state of affairs in the majority of clinical trials. One of the reasons that the medical community continues to use the pain assessment scale is its simplicity of use. The development of a multidimensional tool that combines the different dimensions of pain (sensory, functional, psychological, ...) in a simple and rapid manner could improve pain assessment and treatment recommendations.

1.2.3 Longitudinal evolution of pain

Acute pain is the awareness of a noxious signal from a recently damaged tissue. Its intensity varies according to inflammatory processes, tissue healing and movement. The duration of onset of acute pain is one of its main characteristics (duration less than 3 months). Acute pain is part of everyday life and is generally of little concern because everyone knows from experience how to interpret its cause and predict its outcome. This pain can be excruciating, but it is better tolerated and causes less suffering because it has a limited duration and may be necessary to achieve a worthwhile goal, such as a good sports performance or childbirth. In contrast, in the case of chronic pain, the persistent sense of helplessness and suffering and the long-term impact on daily life generate suffering in patients [34].
Currently, the literature recognizes that chronic pain differs from acute pain in terms of its causes, psychopathology, social significance and impact on brain activity [35, 36]. It has also been shown that the correlations between the different dimensions of pain also become stronger over time, not only in moving from acute to chronic pain, but also during the course of chronic pain [22]. This may be due to the long-term accumulation of bio-psychosocial impacts generated by chronic pain and to alterations in brain mechanisms linking pain to emotion. Therefore, as discussed above, borrowing the principles of assessment and management of acute pain and palliative care pain to treat chronic pain is not adequate. The assessment of chronic pain must take into account the evolving/varying nature of the structure comprising the different dimensions of pain.

1.3 Thesis objectives

In this thesis, we will focus on chronic pain after spinal surgery. Chronic Pain After Spinal Surgery (CPASS), recently included in the 11th revision of the International Classification of Diseases (ICD-11), is defined as chronic pain that occurs after surgery and persists beyond the healing process [37]. This type of chronic pain affects approximately 20% of patients undergoing lumbar spine surgery. The management of these patients represents a major public health problem [38]. As a subclassification of chronic pain, chronic pain after spine surgery also represents the characteristics discussed above: the multidimensionality of its assessment and the longitudinal nature of the association between its dimensions. These two characteristics are not currently taken into account by pain assessment tools. In order to remedy this, we will develop new longitudinal multidimensional statistical analysis techniques to achieve the following objective:
To develop a statistical model allowing to obtain multidimensional scores of pain evaluation whose components evolve with time and according to the various heterogeneous subpopulations. This type of model will allow us, at a given moment, to generate a score for each patient based on a weighted combination of several criteria. Due to the fact that the mechanisms and impact of pain differs from one patient to another. In addition, different combinations are used for different subpopulations and for different points in the patient journey.

1.4 Thesis layout

We begin in chapter 2 by introducing the various statistical concepts and methods proposed in the literature for dealing with heterogeneous longitudinal data. The methods discussed include mixed effects models [3], variable coefficient models [39] and longitudinal factor analysis models [39]. We will also introduce the notion of mixture of models which is a method that allows us to take into account the heterogeneity of the data [40].
Chapter 3 will consist of an application of a finite mixture of mixed effects models on a database of patients with chronic pain after spinal surgery. This chapter will illustrate the usefulness of mixture models in longitudinal data analysis. Our objective in this chapter will be to extract clusters of patients based on the impact of functional disability, pain intensity, and psychological distress on their quality of life and to identify the sociodemographic, behavioral, and clinical characteristics associated with these clusters. The analysis in this study was performed on a dataset of 200 patients from the French multicenter observational study PREDIBACK including patients with chronic pain after spine surgery with a one-year follow-up. The identification of these classes will allow a better understanding of how patients' quality of life will be impacted by the different dimensions of chronic pain and also an opening towards a personalized assessment of pain and quality of life. In this thesis, this application will allow us to introduce mixed effects models as well as to show the heterogeneity of the population of chronic pain patients in relation to how they are impacted by the functional, sensory and psychological components of chronic pain.
The purpose of chapter 4 is to present the statistical development of a time-dependent mixture of non-parametric variable coefficient models including random effects represented by Gaussian processes. This development stems from an applicative necessity due to the

fact that variable coefficient models do not take into account data heterogeneity or intra-subject correlation. This corresponds to a major bias due to the non-inclusion of latent variables corresponding to the random effects and latent classes modeling the homogeneous sub-samples. We will present a specific backfitting procedure to estimate our model. A cross-validation method for the selection of smoothing parameters will also be presented. The proposed model will be evaluated on simulated and real data, showing that this algorithm gathers subjects in homogeneous groups and allows to obtain better estimates of intra-subject and between-subject effects. This model can be considered as a non-parametric version of the mixture of mixed-effects models applied in Chapter 2. We will apply this model on 2 datasets to illustrate its usefulness. First, the model will be applied on the PAQUID database with 500 elderly subjects, in order to estimate the evolution over time of the effect of visual memory and verbal memory on the cognitive deficit. Finally, we will apply the developed model on the PREDIBACK database in order to validate the clusters obtained in chapter 2 and to evaluate their evolution in time.
In chapter 5, we will propose a new method based on a mixture model of longitudinal factor analysis. In a general framework, this model will allow for the description of several dependent longitudinal variables in terms of several independent longitudinal variables through a time-varying factor analysis structure. The model also includes mixing to allow factor structures to vary across individuals. We apply the model to the PREDIBACK database where patients are assessed by several questionnaires measuring several aspects such as pain intensity, quality of life, functional capacity, depression, anxiety, catastrophizing and others. Our objective is to reduce all these dimensions into a much smaller number of dimensions that summarize them into latent factors representing more global concepts, taking into account the evolution of pain over time and the difference between individuals of the links between its dimensions.
In chapter 6, we will discuss the work of this thesis, present future research and prospects, and conclude with a general conclusion of the thesis.
Chapter 3 is the subject of a published paper in the *Journal of Clinical Medicine*. Chapter 4 is the subject of an article submitted for publication. Chapter 5 is an ongoing work that will be submitted for publication. In addition, in the context of my activity as a statistician at the Poitiers University Hospital, I had the opportunity to co-author 8 published clinical articles (including 1 as first author available in the appendix) and 4 submitted manuscripts.

Chapitre 2

Background

2.1 Statistical modeling of longitudinal data

Several statistical techniques have been developed in order to model longitudinal data. These techniques include parametric models such as mixed effect models [2, 42], semi-parametric models such as generalized estimating equations, and non-parametric models such as varying-coefficient models [39].
Since the models proposed in this thesis are extensions of the mixed effects model and the variable coefficient model, we will start by introducing these two models in this chapter. Next, we will introduce the mixture models used to analyze heterogeneous data by modeling each latent subpopulation using a different parameters/functions. We will finish by introducing the factor analysis framework used to model data with several outcome variables (multivariate analysis).

2.1.1 Parametric modeling: Mixed effects models

2.1.1.1 Model specification and interpretation

Mixed effects models [3] are one of the most widely used statistical methods to analyze longitudinal data both from a theoretical and practical point of view. Aside from estimating the effects of several fixed/repeated independent variables, these models also allow regression and analysis of variance/covariance structure of these effects. Let us assume that we have n subjects with subject i having n_i repeated measures of an outcome variable Y which we want to explain using p repeatedly measured covariates X. Mixed effects models can be written as the following:

$$Y_{ij} = X_{ij}\beta^T + Z_{ij}\mu_i^T + \epsilon_{ij}, \tag{2.1}$$

where Y_{ij} is the outcome of subject i at the repeated measure j ($i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, n_i\}$). The vector $X_{ij} \in \mathbb{R}^p$ represents the covariates associated with the fixed effects $\beta \in \mathbb{R}^p$. Fixed effects are effects with are constant across subjects (*i.e.* we assume that the change X cause on Y is the same for all subjects). the vector Z_{ij} of length q represents the covariates associated with the random effects $\mu_i \in \mathbb{R}^q$. The random effects

are observations of Gaussian distribution with mean 0 and a variance/covariance matrix $G \in \mathcal{M}_q(\mathbb{R})$. ϵ_{ij} represents the error terms for subject i at time j. The error terms are generally supposed to be independent observations of a Gaussian distribution with mean 0 and a variance matrix $R = \sigma^2 I_N$, where $N = \sum_{i=1}^{n} n_i$ is the total number of observations. This means that the outcome variable Y_i follows a multivariate Gaussian distribution $\mathcal{N}(X_i\beta, Z_i G Z_i^T + R)$. We also assume that the error terms are independent from the random effects.
The mixed effects model is interpreted through its coefficients β and the variance/covariance matrix G. The estimation of β requires the estimation of the matrices G and R for whom the structure must be prespecified.
One of the significant advantages of the mixed effect model over the classical linear model is that random errors and random effects variance/covariance components allow to model both heterogeneous variances and correlation between observations through the specification of the variance/covariance matrix structures for the random effects μ and the errors ϵ.
Mixed models naturally extend the classical linear models by allowing for the addition of random effects which are random realizations of the population effects distribution. These random realization are pooled to estimate the "pooled mean" effect which represents the global population fixed effect. When the variance/covariance structure is incorrectly specified, but the covariates and random effects are independent, then parameter estimates and their confidence intervals are valid [43]. In this case, the mixed effects model and the classical regression model yield similar fixed effect estimates. On the other hand, when the covariates and random effects are dependent and the variance/covariance structure is incorrectly specified, then high bias could be observed [44]. To illustrate this, we simulated data from 5 subjects with an average of 20 observations per subject. We start by simulating the random intercepts and slopes of the subjects with a Gaussian distribution (mean 0 and standard deviation 4 for the intercept and mean -0.5 and standard deviation 0.2 for the slope). Then, we simulate the predictor variable x for each subject "i" with a Gaussian distribution with mean the intercept of subject "i" and standard deviation 1. Finally, we take :

$$y_i = intercept_i + xi * slope_i + \epsilon$$

Where ϵ are *iid* errors generated from a standard Gaussian distribution. With this simulation, we wanted to have random effects that are inverse to the effect estimated with a classical regression model which shows that the classical regression model can lead to contradictory results and does not allow to estimate the true effect in case of repeated measurements. Knowing that the true slope is -0.5, the classical regression model estimates a slope of 0.32 while the mixed effect model yields a slope of -0.42 which gives completely different interpretations of the relationship between x and y. The results of this simulation are illustrated in Figure 2.1. When intra-subject correlations are present in unbalanced data, a model with no random effects would be underspecified and therefore, would yield wrong inferences. Model selection criteria could be used to select the optimal variance/covariance structure (*e.g.* whether to include random effects in the model or not) [45–48].

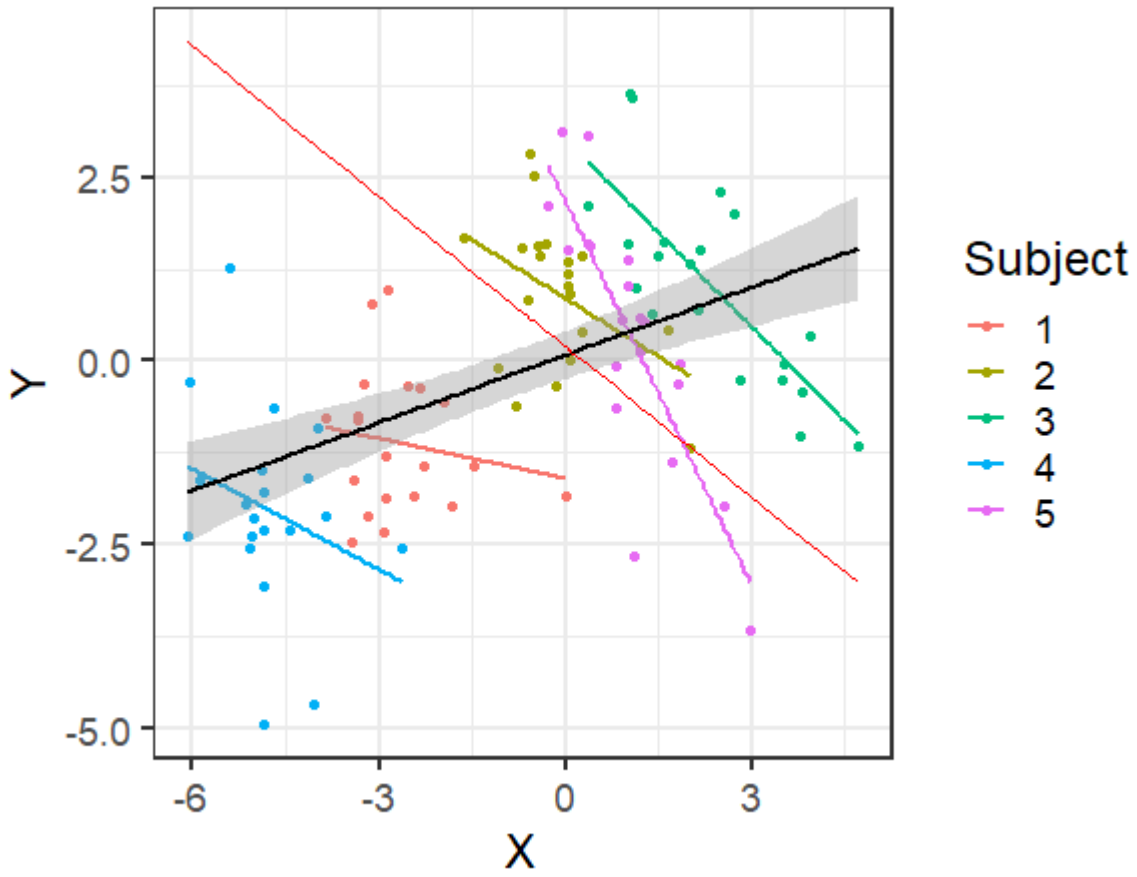

FIGURE 2.1 – each color represents a subject. The black line is the fitted line using a simple linear model with no random effects and the red line is fitted using a mixed effects model. We can clearly see that the intercept and slope are completely different between the two estimation methods. The simple linear model estimates the global effects based on the full sample while the mixed effects model estimates the effects by "pooling" subject effects.

2.1.1.2 Model parameters estimation

Several approaches can be used to obtain the covariance matrix estimates including maximum likelihood (ML) or restricted maximum likelihood (REML).
The log-likelihood function for the ML approach can be written as the following:

$$\mathcal{LL}_{ML} = \sum_{i=1}^{n} \left\{ -\frac{1}{2} \log(|Z_i G Z_i^T + R|) - \frac{1}{2}(Y_i - X_i\beta)^T (Z_i G Z_i^T + R)^{-1} (Y_i - X_i\beta) \right\} - \frac{N}{2} \log(2\pi).$$

This maximization problem does not have an analytical solution. Therefore, the model is estimated using numerical and approximation methods including Expectation-Maximization (EM) and Newton-Raphson algorithms [49] or by combining the penalized least squares algorithm and a nonlinear optimization algorithm such as Bound Optimization BY Quadratic Approximation (BOBYQA) algorithm [50].

2.1.2 Non-parametric modeling: Varying-coefficient models

Varying-coefficient models [39] are non-parametric models which are widely used to estimate the time-varying impact of several longitudinal explicative variables on a longitudinal outcome. In this model, the time-varying effects are assumed to be smooth functions of time.
Keeping the same notations as above, the time-varying coefficient models can be written as follows:

$$Y_i(t_{ij}) = X_i(t_{ij})^T \beta(t_{ij}) + \epsilon_i(t_{ij}), \tag{2.2}$$

where $Y_i(t_{ij})$ is the outcome of subject i at time t_{ij} ($i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, n_i\}$). The vector $X_i(t_{ij}) \in \mathbb{R}^p$ represents the time-varying covariates associated with the smooth time-function effects $\beta(t_{ij})$. The error terms $\epsilon_i(t_{ij})$ are independent observations of a Gaussian process $\epsilon(t)$ with mean 0 and a variance function $\sigma(t)$.
One of arguably the most important advantages of the time-varying coefficient model is its flexibility in modeling the effects non-parametrically which can not be achieved using mixed effects models. Time varying coefficients can be achieved in mixed models only by adding a prespecified function of time as a new covariate which does not provide as much flexibility as non-parametric methods. On the other hand the mixed effects model allows the estimation of the variance/covariance structure which can not be estimated in the classical varying-coefficient model.
Several approaches have been developed to estimate the time-varying coefficient $\beta(t)$ including kernel local polynomial smoothing [51–53], smoothing splines [39, 53, 54] and polynomial splines [55, 56]. The variable coefficient models, such as they are defined, are locally linear models. It is more reasonable to use the kernel smoothing method to estimate them. In what follows, we will introduce the kernel local polynomial smoothing method since it is the method used in this thesis.

2.1.2.1 Model parameters estimation

The estimation of the time-varying coefficient model in equation (2.2) using the kernel local linear smoothing corresponds to the maximization of the following local log-likelihood

function:

$$\sum_{i=1}^{N}\sum_{c=1}^{C}\log[\phi(Y_i|\mathrm{X}_i^T\beta,\sigma)]k_h(t_i-t), \tag{2.3}$$

where ϕ is a density function (generally taken as the Gaussian density function). The kernel K_h is a positive symmetric real-valued function, which has an integral of 1 over $\mathbb{R}$. We obtain the estimates of $\hat{\beta}(t)$ and $\hat{\sigma}(t)$ by interpolating the local estimates $\hat{\beta}$ and $\hat{\sigma}$ obtained by maximizing the local log-likelihood at (2.3). In case the density function ϕ is normal, $\hat{\beta}(t)$ and $\hat{\sigma}(t)$ have explicit solutions given by:

$$\hat{\beta}(t) = (S^TWS)^{-1}S^TWY,$$
$$\hat{\sigma}(t) = (Y - X^T\hat{\beta})^TW(Y - X^T\hat{\beta})/tr(W)$$

where $S = X^T$, $W = diagw_1, \ldots, w_n$ with $w_i = k_h(t_i - t)$. The quantity $tr(W)$ is the trace of the diagonal matrix W.

On the other hand, when ϕ is not normal (generalized varying-coefficient models as in [57–60]), the model is estimated using the usual numerical methods such as Newton-Raphson algorithm [61].

One of the limitations of this model is that it does not take into account the intra-subject correlation which is a usually present in longitudinal data since each subject has multiple observations (repeated measurements). This correlation needs to be taken into account in the analysis in order to achieve valid inferences. Several authors proposed varying-coefficients models which take into account intra-subject dependencies. Early attempts [53, 62] consisted of estimating the model by pooling data from all subjects and indirectly considering intra-subject correlation by estimating the optimal bandwidth h of the used kernel function, using leave-one-subject-out cross validation which was shown to be better when data are correlated [63]. More recently, other authors proposed varying-coefficient models for longitudinal data with repeated measurements, either indirectly by including covariances in the error term of the model [53, 64, 65] or directly by including random effects in the model (2.2) [66–68] similar to the following model:

$$Y_i(t_{ij}) = X_i(t_{ij})^T\beta(t_{ij}) + Z_{ij}^T\mu_i + \epsilon_i(t_{ij}), \tag{2.4}$$

where Z_{ij}^T is the vector of covariates associated with the random effects μ_i. The random effects μ_i are assumed to be independent from the error terms $\epsilon_i(t_{ij})$ and are assumed to follow a Gaussian distribution with a mean 0 and covariance matrix D which models the correlations between the random effects. The kernel estimation method uses local observations in the estimation procedure for the fixed effects but the random effects are represented by global invariant random coefficients. This results in models which ignore the local variations between individuals. Therefore, these methods might lead invalid inferences and loss of efficiency. One of the solutions proposed in the literature is to include time-varying random effects represented by random processes [69, 70]. The aforementioned model can be written as follows:

$$Y_i(t_{ij}) = X_i^T(t_{ij})\beta(t_{ij}) + Z_i(t_{ij})\mu_i(t) + e_i(t_{ij}), \tag{2.5}$$

where $\mu_i(t)$ are the random effects function that characterizes individual variations (subject-effect) from the fixed effects $\beta(t_{ij})$. In addition, $\mu_i(t)$ are independent observations

of stochastic processes $\mu(t)$ indexed by the time variable t. It is assumed that $\mu(t)$ is a Gaussian process with mean 0 and covariance function $\rho(s,t)$. The random effects component help us incorporate the intra-subject feature of longitudinal data and the covariance structure. It was shown that the model (2.5) proposed in [70] yields less biased estimates of the true effects compared to the models with correlated error terms [53] even more when intra-subject correlation is strong.

The model (2.5) can be estimated using a two step procedure similar to the one proposed by Fan & Zhang [71]. This estimation procedure consists of getting raw local estimates at each time point using the classical mixed effects model and then applying a smoothing procedure on these local coefficients. Wu & Liang proposed a backfitting procedure (see [72] and [39]) which simplifies computations and helps with multicollinearity. The backfitting is an iterative method consisting of fitting the coefficient associated with each dependant variable separately at each iteration and then replacing the coefficient in the model by its estimate. For model (2.5), the backfitting algorithm proposed by Wu & Liang for the case when X and Z are equal, is the following:

1. Input: outcome vector Y, covariate matrices X and Z, and the time variable t.
2. Initiate the values of $\beta(t) = (\beta_0(t), \ldots, \beta_p(t))$ and $\mu_i(t) = (\mu_{i0}(t), \ldots, \mu_{ip}(t))$ where p is the number of covariates.
3. For $l = 0, \ldots, p$:
   (a) For all $i \in \{1, \ldots, n\}$ and for all $j \in \{1, \ldots, n_i\}$, let $\widetilde{Y}_i^{(l)}(t_{ij}) = Y_i(t_{ij}) - X_i^{(-l)T}(t_{ij})\beta^{(-l)}(t_{ij}) + Z_i^{(-l)}(t_{ij})\mu_i^{(-l)}(t)$, where the index $(-l)$ represents the vector without its l^{th} component (*e.g.* $\beta^{(-l)}(t) = (\beta_0(t), \ldots, \beta_{l-1}(t), \beta_{l+1}(t), \ldots, \beta_p(t))$).
   (b) Fit the model:
   $$\widetilde{Y}_i^{(l)}(t_{ij}) = X_i^{(l)T}(t_{ij})\beta_l(t_{ij}) + Z_i^{(l)}(t_{ij})\mu_{li}(t) + e_i^{(l)}(t_{ij}), \tag{2.6}$$
   to get $\hat{\beta}_l(t)$ and $\hat{\mu}_{li}(t)$, which are respectively the estimates of the time-varying fixed effect and random effect of the l^{th} dependant variable.
   (c) Replace $\beta_l(t)$ and $\mu_{li}(t)$ by $\hat{\beta}_l(t)$ and $\hat{\mu}_{li}(t)$ respectively.
4. Repeat step 3 until convergence.
5. Output: The final estimates of the time-varying fixed effects and random effects $\hat{\beta}(t)$ and $\hat{\mu}(t)$.

In step 3.*b* in the algorithm above, the model (2.6) is estimated using the two-step procedure discussed before by fitting mixed linear models locally and then using a smoothing method to obtain the time-varying effects.

The model proposed by Wu and Liang [70] above is able to estimate both inter and intra-subject effects using time-varying smooth functions and gaussian random processes respectively.

Apart from the inability of the classical varying-coefficient models to model intra-subject effects, another limitation is their inability to model heterogeneous data gathered from different subpopulations. This limitation can be appropriately managed by estimating

different effects for each group of subjects depending on the likelihood of a given model in their subpopulation. One solution called mixture of regression models was proposed for the context of classical regression models by DeVeaux [40]. Mixture of regressions generate a different model for each homogeneous subpopulation, which allows them to take into account the heterogeneity of data.
In the next section, we will discuss the mixture of regressions model, its parameters and its estimation procedure. We will also discuss the recent extensions of this model that are relevant to this thesis.

2.2 Statistical modeling of heterogeneous data

the expression "Heterogenous data" has several meanings in the statistical community. The industrial use of the term refers to data from different sources (*e.g.* medical records, imaging, quastionnaires, ...). In statistics, the term heterogeneous data refers to samples that are drawn from several populations with different distributions. Heterogeneous populations are very common in real life. For example, in clinical data, patients are typically a very heterogeneous population since they differ regarding their multivariate distribution including demographics, diagnosis, and history of the disease. This can be especially witnessed in the differences in findings in the medical literature and the development of methods to handle heterogeneity in meta-analysis [73].
In the next subsection, we will describe mixture of regression models which can be used to model heterogeneity in the relationship between a predictor and an outcome.

2.2.1 Mixture of models

Identifying individual differences in the relationship between a predictor x and an outcome y, has become increasingly important in areas of the health and social sciences. Understanding these differences in effects is important because it allows us to answer the question "for whom does x predict y ?" instead of the classical regression hypothesis "does x predict y ?". To understand the complex ways in which psychological, social and clinical factors impact the relationship between health outcomes, we need to consider the explicit empirical effect-variability among patient subpopulations.
Mixture of regression models are an exploratory approach from the finite mixture models framework, which search for evidence of heterogeneity in the effects of one or several predictors on an outcome. Unlike the use of statistical interactions (which test whether the effects of a predictor on an outcome vary between the categories of a third measured variable), mixture of regressions ties to find groups of respondents who differ in the effects of predictors on the studied outcome. The method was first introduced in the economics literature in the form of a general switching regression models by Quandt and Ramsey [74, 75]. Their technique was based on a moment generating function to estimate the parameters. These methods have been since applied in medicine, genetics, agriculture [76, 77]. These models work by empirically identifying subgroups that are impacted differently by a predictor, and have been successful at finding subpopulation effects that other methods could not find [78]. Those differences in effects can be identified using latent categorical variables (unobserved variables), which represent the subpopulation associated

with the different effects.

2.2.1.1 Model formulation and estimation

The mixture of regressions model considers the same setting as the classical regression model. Let n be the number of subjects. Let Y_i and $X_i = (X_{1i}, \ldots, X_{pi})$ be the observed outcome variable and the predictors for subject i ($i \in \{1, \ldots, N\}$) respectively. The mixture of regressions model can be written as follows:

$$Y_i = \sum_{c=1}^{C} \mathbb{1}_{\{z_i=c\}}(X_i^T \beta_c + \epsilon_{ci}), \tag{2.7}$$

where C is the number of clusters/subpopulations. The quantity indicating the cluster membership for individual z_i, for individual i, is an observation from random categorical latent variable z with a set of possible values $\{1, \ldots, C\}$ with probabilities $\{\pi_1, \ldots, \pi_C\}$ with $\sum_{c=1}^{C} \pi_c = 1$. The coefficients β_c represent the effects of X on Y for the individuals in cluster c. The error terms ϵ_c are normally distributed with a zero mean and a standard deviation σ_c. The model (2.7) models with a different regression equation each cluster. The model (2.7) can also be formulated using in density functions as the following:

$$\phi(Y|\theta, X) = \sum_{c=1}^{C} \pi_c \phi_c(Y|\theta_c, X),$$

where $\theta = (\theta_1, \ldots, \theta_C)$ is the parameters set including the mixture proportions $(\pi_1, \ldots, \pi_C)$, the regression coefficients $\beta_c = (\beta_{0c}, \ldots, \beta_{pc})$ for each mixture component c in $\{1, \ldots, C\}$, and the standard deviations of the error terms $(\sigma_1, \ldots, \sigma_C)$. The function ϕ_c denotes the density of a Gaussian distribution with a mean $X_i^T \beta_c$ and a variance σ_c^2. The log-likelihood of the model is therefore given by:

$$\mathcal{LL}(\theta|X, Y) = \sum_{i=1}^{N} \log(\sum_{c=1}^{C} \pi_c \phi_c(Y_i|\theta_c, X_i)),$$

The model 2.7 can be estimated using the Expectation Maximization (EM) [79] algorithm as in [40]. The EM algorithm can be applied to a multitude of problems with unobserved data (the variable z in model 2.7 is not observed) where the estimates are obtained using maximum likelihood. This type of problems is called modeling with incomplete data since only X and Y are observed in the data consisting of (X,Y,Z). The EM algorithm is an iterative algorithm where at each iteration we alternate between the E-step, which calculates the expectation of the log-likelihood for the parameters and the M-step, which maximizes the expected log-likelihood calculated in the E-step. The EM algorithm for estimating the mixture of regression models can be written as follows:
Let $\theta^{(l)}$ represent the estimates of the model parameters at iteration l.

1. initialize the parameters $\theta^{(0)}$.
2. iterate the following two steps until convergence:

3. *E-step*: Calculate the expectation of the complete-data log-likelihood conditional on the current parameters estimates:

$$Q(\theta, \theta^{(l)}) = \sum_{i=1}^{N} \sum_{c=1}^{C} r_{ic}^{(l)} \phi_c(Y_i|\theta_c, X_i),$$

where

$$r_{ic}^{(l)} = \frac{\pi_c^{(l)} \phi_c(Y_i|\theta_c^{(l)}, X_i)}{\sum_{k=1}^{C} \pi_k^{(l)} \phi_k(Y_i|\theta_k^{(l)}, X_i)}$$

The quantity r_{ic} is referred to as the posterior probability of subject i belonging to the component c.

4. *M-step*: Maximize the function $Q(\theta, \theta^{(l)})$ in θ to get $\theta^{(l+1)}$. The maximization is equivalent to performing a weighted regression with weights $r_{ic}^{(l)}$.

Therefore, when Y is continuous, the function Q in the M-step can be explicitly maximized and the maximization gives the following updates for each mixture component c:

$$\hat{\pi}_c^{(l+1)} = \frac{\sum_{i=1}^{N} r_{ic}^{(l)}}{N}$$

$$\hat{\beta}_c^{(l+1)} = (X^T R_c X)^{-1} X^T W_c Y$$

where $R_c = diag\ (r_{1c}^{(l)}, \ldots, r_{Nc}^{(l)})$ is the diagonal matrix containing the posterior mixture probabilities for each subject.

$$\sigma_c^{2(l+1)} = \frac{\sum_{i=1}^{N} r_{ic}^{(l)} (Y_i - X_i^T \hat{\beta}_c^{(l+1)})^2}{\sum_{i=1}^{N} r_{ic}^{(l)}}$$

2.2.1.2 Some extensions of the mixture of linear regression models

Since in this work we focus on the modeling of longitudinal data, we will discuss some extensions proposed to model heterogeneous longitudinal data. On of the proposed models is the mixture of mixed effects/multilevel models (see [80–83]). This model attempts to resolve the intra-correlation problem in longitudinal data by mixing several mixed effects models instead of a classical regression model. The mixture of mixed effects model can be formulated as follows:

$$Y_{ij} = \sum_{c=1}^{C} \mathbb{1}_{\{z_i=c\}} (X_{ij}^T \beta_c + Z_{ij}^T \mu_i + \epsilon_{ci}). \tag{2.8}$$

As can be seen, the model is written as a mixture of the mixed effect models given in equation (2.1). The model (2.8) is estimated by maximizing the likelihood either by using the EM algorithm or by directly maximizing the likelihood using numerical methods such as the Newton-Raphson algorithm. More details on the model estimation can be found in [80]. An R implementation of this model is available in the package `lcmm` [84]. The mixture of mixed effect models was applied to the PREDIBACK study data in order to identify clusters of patients with chronic pain after spine surgery (see chapter 3).

Another extension of the mixture of regression models proposed for longitudinal data proposed by Huang et al. [85] consists of a mixture of varying-coefficient models. This model is described in details in chapter 4.

2.3 Statistical modeling of longitudinal data with multiple outcomes

In social and medical sciences based on patient reported outcomes (*e.g.* questionnaires), the definition of certain variables is problematic. They are often theoretical constructs that cannot be measured directly (latent variables) and whose "existence" is postulated on the basis of abstract reasoning specific to the field of application. For example, quality of life is a construct which can only be evaluated using a set of noisy indirect questions (*e.g.* mobility and self-care) that are theoretically associated with the concept. These theoretical constructs are observed indirectly by noting their influence on measured variables, such as aptitude tests or responses questionnaire. As discussed in the introduction, the health status of a chronic pain patient is a multidimensional latent factor which can only be evaluated by combining several indicators which are currently evaluated separately. In such cases where patients true outcome is not measurable and is multidimensional, how is the variable that represents the effectiveness of a medical procedure defined ?
Several multivariate methods have been developed to address the issue where multiple outcomes are important and are associated with the patient's true outcome [86–88]. The first approach consists of jointly analysing correlated outcomes based on their multivariate joint distribution like in the multivariate anlaysis of variance (MANOVA). The outcomes are simultaneously combined to a vector random variable following a multivariate distribution with a complex covariance structure. The second approach which will be addressed in this thesis is the use factorial variable reduction techniques such as principal component analysis (PCA) and exploratory factor analysis (EFA). This is generally done in a two step procedure where the outcome are first summarized in a single latent variable which explains a certain amount of the information contained in the outcomes. Then this latent factor is used as a global unique outcome and is evaluated using the classical "single outcome" univariate statistical analysis. Other statistical methods where these two approaches are combined have been proposed in the structural equation modeling framework [89–91]. In this framework, both the model explaining the relationship between the measured outcomes and the latent variables and the model explaining the impact of explanatory variables on the latent outcomes are estimated simultaneously as a system of equations.
We will now introduce the EFA model formulation and its estimation.

2.3.1 Factor analysis

Exploratory factor analysis is a variable reduction technique that identifies the underlying factor structure of a set of variables. The difference between EFA and PCA is that PCA generates components that are orthogonal linear combinations that maximize the total variance while EFA generates factors that maximize the joint variance of the variables within the factors.

Mathematical basics

Let n be the number of subjects and p be the number of observed/measured variables. Let X be a matrix (of dimension $n \times p$) that groups the p observed variables. X_j indicates the observed variable j. Let μ be a vector of size p that contains the means $\mathbb{E}(X_j)$ of the p observed variables. Let us consider m latent (non-observable) variables $f_1,\ f_2,\ \ldots\ ,\ f_m$, which we will call common factors ($m << p$). The factor analysis model can be seen as a system of multiple regressions, where the response variable is X_j ($j = 1, \ldots, p$) and the independent variables are f_k ($k = 1, \ldots, m$):

$$\begin{aligned} X_1 &= \mu_1 + \lambda_{11} f_1 + \lambda_{12} f_2 + \cdots + \lambda_{1m} f_m + \epsilon_1, \\ X_2 &= \mu_2 + \lambda_{21} f_1 + \lambda_{22} f_2 + \cdots + \lambda_{2m} f_m + \epsilon_2, \\ &\vdots \\ X_p &= \mu_p + \lambda_{p1} f_1 + \lambda_{p2} f_2 + \cdots + \lambda_{pm} f_m + \epsilon_p, \end{aligned}$$

where ϵ_j are the error terms which follow a Gaussian distribution with mean 0 and variance ψ_j. The regression coefficients λ_{jk} are called "factor loadings".

$$\Lambda = \begin{pmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{11} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{11} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{11} \end{pmatrix} = \text{factor loadings matrix.}$$

In summary, the basic model is similar to a regression model. Each of the response variables X_j is described as a linear function of the common non-observable factors $f_1,\ f_2,\ \ldots\ ,\ f_m$. Thus we have m latent variables that explain the variation of the p observed variables. The model can also be written in matrix format as the following:

$$X = \mu + \Lambda f + \epsilon.$$

Therefore, the model's variance-covariance matrix is the following:

$$\Sigma = \Lambda \Lambda^T + \Psi,$$

where Ψ is equal to:

$$\Psi = \begin{pmatrix} \psi 1 & 0 & \cdots & 0 \\ 0 & \psi 2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \psi p \end{pmatrix},$$

With ψ_j is equal to $\text{Var}(\epsilon_j)$. The model parameters θ to estimate are the factor loadings matrix Λ and the error terms variance matrix Ψ. The number of parameters is therefore equal to $mp + p$.

2.3.1.1 Factor analysis model estimation

Several methods have been proposed to estimate the factor analysis model parameters including algebraic methods such as the principal factors method where the parameters are

estimated using the spectral decomposition of the empirical variance-covariance matrix $\hat{\Sigma}_{X_{cs}}$ (X_{cs} the standardized and centred version of X). In this thesis we will focus on the maximum likelihood estimation of the factor analysis model.

Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ be the empirical mean of the observed variables and let $S = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})^T$ be the empirical variance of the observed variables. Therefore, the log-likelihood function of the model can be written as:

$$\mathcal{LL}(X, \mu, \Sigma) = -\frac{n}{2}(\log|\Sigma| + tr(\Sigma^{-1}S) + (\overline{X} - \mu)^T\Sigma^{-1}(\overline{X} - \mu))$$

The maximum likelihood estimator of the parameter μ is obviously the sample means $\overline{X}$. Therefore the log-likelihood of the model reduces to:

$$\mathcal{LL}(X, \mu, \Sigma) = -\frac{n}{2}(\log|\Sigma| + tr(\Sigma^{-1}S)) \tag{2.9}$$

The derivation of equation (2.9) regarding the model parameters yields to the following estimating equations:

$$\frac{d\mathcal{LL}(\theta)}{d\Lambda} = -\frac{n}{2}(\Sigma^{-1}\Lambda - \Sigma^{-1}S\Sigma^{-1}\Lambda) = 0, \tag{2.10}$$

$$\frac{d\mathcal{LL}(\theta)}{d\Psi} = -\frac{n}{2}diag(\Sigma^{-1} - \Sigma^{-1}S\Sigma^{-1}) = 0. \tag{2.11}$$

To obtain the model parameters, the two equations above need to be solved simultaneously. Obtaining explicit solutions of this maximization problem is difficult thus some authors proposed iterative methods to get the parameters estimates [92, 93]. Among these methods, the EM algorithm offers a simple and efficient solution to estimate the model. The EM algorithm for solving the model can be found in a paper by Rubin and Thayer (1982) [93] or in a paper by Zhao *et al.* [94].
Several extensions of factor analysis models have been proposed in the literature [7, 8, 12, 95]. In this thesis we focus on two extensions, the longitudinal mixed effects factor analysis and the mixture of factor analysis. These two extensions were used as foundations of our proposed time-varying mixture of longitudinal factor analysis model.

2.3.2 Longitudinal factor analysis

As can be seen in the maximum likelihood estimation of the classical factor analysis previously discussed, the observations are assumed to be independent ; however, this independence should not be assumed when dealing with longitudinal repeated-measures data. To address this limitation of the classical factor analysis framework, Roy & Lin (2000) [6] proposed a factor analysis model where a random intercept is allowed. We use the same notations as those used in the previous section. the model proposed by Roy and Lin can be written as follows:

$$X_{ijk} = \mu_j + \lambda_j f_{ik} + \eta_{ij} + \epsilon_{ijk}, \tag{2.12}$$

where $i \in \{1, \ldots, n\}$ represents the subject i data, the index j represents the measured variable j and k is the measure obtained at time k for subject i. the new component η_{ij} is the random intercept used to model the intra-subject correlations in X_{ijk} over time if f_{ik} were observed. The random intercepts η_{ij} are supposed to follow a Gaussian distribution with mean 0 and a covariance matrix D_j. The random intercepts are also assumed to be independent between subjects. In the model proposed by Roy and Lin, only one latent factor f is considered. The authors also proposed to use a mixed effects model to estimate the impact of given explanatory variables on the latent factor f. Let T and Z be the matrices of the explanatory variables of fixed effects and random effects respectively. The mixed effects model they proposed is the following:

$$f_{ik} = T_{ik}^T \beta + Z_{ik}^T \xi_i + \omega_{ik}, \tag{2.13}$$

where β represents the fixed effects and ξ_i is a vector representing the subject-specific random effects which follow a Gaussian distribution with mean 0 and a covariance matrix G. The random observation ω_{ik} represent the error term which also follows a Gaussian distribution with mean 0 and variance σ^2. The models comprising the equations (2.12) and (2.13) are estimated simultaneously using the EM algorithm to maximize its likelihood function.

One of the limitations of the model above is that only one latent factor can be considered. In some situations several latent factors can be needed to adequately incorporate the information from a high number of outcomes. To address this limitation An *et al.* (2013) [7] proposed a factor analysis mixed model similar to Roy and Lin model where the number of latent factors is not restricted to one.

The two equations representing the factor analysis mixed model are:

$$X_{ijk} = \mu_j + \Lambda_j f_{ik} + \eta_{ij} + \epsilon_{ijk}, \tag{2.14}$$

$$f_{ik} = T_{ik}^T \beta + Z_{ik}^T \xi_i + \omega_{ik}, \tag{2.15}$$

where the factor loadings, for outcome j, $\Lambda_j = (\lambda_{j1}, \ldots, \lambda_{jm})$ is now a vector of length m and the latent factors are $f_{ik} = (f_{ik1}, \ldots, f_{ikm})$ is a vector of factor scores for each latent factor. The model by An *et al.* is also estimated by maximizing the model log-likelihood function using the EM algorithm (See [7] for more details). One of the restrictions of the previous model is the use of a fixed loading matrix over time (time-invariance) and between subject (group invariance). These assumptions are not always verified in real world applications. In the literature, measurement invariance between groups is tested on known groups (*e.g.* testing the difference across countries/cultures of the g-factor intelligence model structure) [96, 97]. In practice, differences in the factor structure can be associated with both known and unknown variables. To address this some authors proposed the mixture of factor analysers which will be discussed in chapter 5.

Chapitre 3

Application of mixture of mixed effects models for stratifying the health-related quality of life of chronic back and leg post-operative pain patients

Résumé du chapitre en français

La douleur chronique après une chirurgie du rachis a un impact dramatique sur la qualité de vie des patients, comme le montrent les outils d'évaluation de la Qualité de Vie Liée à la Santé (QVLS). Cependant, l'importance de la capacité fonctionnelle, de la perception de la douleur et de l'état psychologique dans la qualité de vie des patients peut varier considérablement d'un sujet à l'autre. Notre objectif était de clusteriser les patients en fonction de l'impact des différentes dimensions des douleurs chroniques sur la QVLS dans un échantillon de patients avec des douleurs chroniques après une chirurgie du rachis. Le deuxième objectif était d'identifier les facteurs associés aux groupes de patients identifiés précédemment. Deux groupes ont été identifiées en utilisant un mélange de modèles à effets mixtes à partir d'un ensemble de données cliniques de 200 patients inscrits à "PREDIBACK", une étude prospective observationnelle multicentrique incluant des patients avec des douleurs chroniques après une chirurgie du rachis, avec un suivi d'un an. Nous avons observé que la QVLS était plus impactée par l'incapacité fonctionnelle pour les patients du premier groupe (n=136) et par la perception de la douleur pour les patients du deuxième groupe (n=62). Les hommes qui perçoivent leur travail comme physique étaient plus affectés par le handicap que par l'intensité de la douleur. Un niveau d'éducation plus faible, l'absence de stratégies d'adaptation et une intensité de la douleur plus élevée étaient significativement associés à une QVLS plus affectée par la perception de la douleur. L'identification de ces groupes permet de mieux comprendre les dimensions de la QVLS et ouvre la voie à une évaluation optimisée de la qualité de vie liée à la santé et à une gestion

personnalisée de la douleur.

3.1 Introduction

Persistent Spinal Pain Syndrome Type 2 (PSPS-T2) [98], previously known as Failed Back Surgery Syndrome or chronic pain after spinal surgery, has been recently identified in the International classification of diseases 11th revision (ICD-11) [37]. PSPS-T2 is characterized as chronic pain developing after a spinal surgical procedure and persistent beyond the healing process [98]. With a prevalence of approximately 20% of patients who undergone spine surgery, PSPS-T2 represents a major public health and a financial burden [38]. Besides pain perception, PSPS-T2 patients also present psychological distress and functional disability, which lead to a decrease in Health Related Quality of Life (HRQoL) [99]. The weighted impact of pain, psychological distress and functional disability on HRQoL have yet to be determined to more accurately characterize PSPS-T2 patients, and finally provide tailored optimal care. Previous research, investigating relationship between quality of life and clinical outcomes in different pathologies, found that pain, psychological distress and functional disability were correlated with HRQoL [100–103]. More specifically, Kovacs et al. [104] investigated the correlation between pain, functional disability and HRQoL before and after a follow-up period of 14-days in 195 patients suffering from low back pain. The authors have shown that HRQoL was significantly negatively correlated with pain intensity and functional disability (coefficient of correlation r ranging from -0.347 to -0.672). Furthermore, in a cross sectional study on 200 patients with chronic pain, Rapti et al. [105] have found significant negative correlations between HRQoL and pain intensity (r=-0.65), and depression score (r=-0.63). All these findings have been however obtained from global correlation. This approach does not allow defining the potentially specific sub-populations related to the impact of pain intensity, functional disability and psychological distress on HRQoL. Clustering patients based on the impact of each chronic pain component on HRQoL could be performed to identify and to characterize specific PSPS-T2 patient's classes. These findings will help improve patients' HRQoL all along their care pathway by managing pain with targeted therapies.

The identification of clusters can be performed using finite mixture models [80, 106, 107]. These models allow providing two or more distinct homogeneous latent classes of individuals based on the associations between different variables. For example, Chen et al. [106] used finite mixture models in order to extract the trajectories of pain over a 5-year follow-up in patients with low back pain. The authors found four pain trajectories profiles: no or occasional mild pain, persistent mild pain, fluctuating pain, and persistent severe pain. By characterizing the population, they notably reported that level of education (<16 years), restricted work or unemployment, pain intensity, disability, pain duration (> 3 years), catastrophizing, anxiety and depression were associated with pain trajectories. While they determined clusters based on trajectories, this study was not intended to determine clusters according to pain component factors (pain intensity, functional disability and psychological distress) related to quality of life. Thereby, clustering and characterization of chronic low back pain patients regarding association between different pain components have yet to be determined, especially in PSPS-T2 population.

The main aim of this real-life prospective multicentric 12-month follow-up study was to identify subgroups based on the impact of pain intensity, functional disability and

psychological distress on HRQoL in 200 PSPS-T2 patients. Our secondary objective was to characterize classes with clinical, psychological, and social factors.

3.2 Methods

3.2.1 Data collection and PREDIBACK study description

The data used for the purpose of this research consisted of PSPS-T2 patients included in the prospective, multicenter, observational PREDIBACK study. The primary objective of the PREDIBACK study was to characterize PSPS-T2 patients based on clinical, psychological and sociodemographic factors. Two hundred patients were consecutively recruited in the pain management department of 5 French centers (Angoulême, Bressuire, La Rochelle, Niort and Poitiers) and monitored at baseline, 3-month, 6-month, 9-month and 12-month follow-ups. Patient recruitment started in January 2017 and was completed in March 2018. The study was approved by the ANSM (2016-A01144-47) and the Ethics Committee (CPP Ouest III) and declared at https://clinicaltrials.gov/ct2/show/NCT02964130.

3.2.2 Patient Selection

3.2.2.1 Inclusion criteria

PSPS-T2 patients were identified at each site through standard clinical practice. To be eligible for this study, patients had to have had at least one spinal surgery, post-operative leg and/or low back pain for at least six months, and an average global pain score greater than or equal to 4 as measured by the Numeric Pain Rating Scale (NPRS). All the patients gave their informed consent before enrolment.

3.2.2.2 Non-inclusion criteria

Patients with one or several of the following criteria were excluded from the study: patient is or has been treated with spinal cord, subcutaneous or peripheral nerve stimulation or an intrathecal drug delivery system, has life expectancy of less than 12 months beyond study enrollment, patient is unable to undergo study assessments or to complete questionnaires independently, is a member of a vulnerable population, and investigator suspects substance abuse which might confound the study results.

3.2.3 Demographic variables encoding

Patient professional activity was encoded as follows: patients were considered as (i) active when they were involved in professional activity at baseline assessment, (ii) inactive when they were retired, without any professional activity, or on sick leave, disability and long-term sick leave. There were no patients that are still students. Two patients were housewives and were considered as active. Furthermore, to identify active patients with arduous working conditions, they were asked to characterize their work as "physically demanding" or not.

3.2.4 Measurement methods and their psychometric properties

3.2.4.1 Health Related Quality of Life

The French version of the EuroQuol 5-Dimensions 5-Level questionnaire (EQ-5D-5L)[108] was used to measure HRQoL. The questionnaire comprises five items including pain intensity, mobility, self-care, daily activities and psychological state (Anxiety or depression). Each item consists of a 5 level Likert scale ranging from "I have no problem" to "I am unable to". The maximum score of 1 indicates the best possible HRQoL. EQ-5D has demonstrated moderate to excellent validity and reliability in a large number of chronic conditions studies [109–111].

3.2.4.2 Pain intensity

Pain intensity was measured using the NPRS [112] which is an 11 point (0-10) unidimensional scale were the patient selects a number that best represent his pain intensity. The 11-point numeric scale ranges from 0 representing "no pain" to 10 representing "the worst pain imaginable".

3.2.4.3 Functional disability

The Oswestry Disability Index (ODI) questionnaire [113] provides a global assessment of pain intensity, functional capacities and specific disability related to acute, subacute or chronic low back pain. It indicates the degree of disability induced by the patient's low back pain. The questionnaire consists of 10 items ranging from 0 to 5 where 0 indicates high ability to perform the task associated with the item and 5 indicates the inability to perform the task. Items included pain intensity, degree of disability for personal care, lifting, walking, sitting, standing, sleeping, sexual life, social life and travelling. The score is expressed as a percentage of disability. The ODI is one of the questionnaires that has been extensively validated by a large number of studies showing an internal consistency ranging from moderate to excellent [114, 115].

3.2.4.4 Depressive disorder

The Hospital Anxiety and Depression Scale (HADS) [116] was used to investigate anxiety and depression symptoms and their severity. The questionnaire comprises 14 items each comprising 4 levels and either representing a symptom of anxiety or depression. The total score ranges from 0 to 24 for each category (depression or anxiety). A score of 11 or above indicates a definite symptomatology. The HAD questionnaire showed high internal consistency in the acute low back pain population and in the general population [117, 118]. HADS has been used as an assessment tool in several studies on chronic back and leg pain [119].

3.2.4.5 Coping strategies

The French adaptation of the Coping Strategies Questionnaire (CSQ) [120] was used to assess the patient self-rated use of cognitive and behavioral strategies to cope with pain. It consists of six subscales for cognitive strategies including pain ignorance (5 items), reinterpretation

(4 items), diversion (5 items), self-encouragement (4 items), catastrophizing (6 items) and praying/hoping (3 items). The praying subscale was not used in this study because it is considered as sensitive data. Each item consists of a Likert scale ranging from 1 "never" to 4 "always" indicating how frequently the strategy is used to cope with pain.

3.2.5 Statistical analysis

Statistical analyses were conducted using the R software (Version 3.6.0, R Foundation for Statistical Computing, Vienna, Austria).
Variables were described by their means (standard deviation) or by their number (percentage) depending on whether the variable was qualitative or quantitative. Missing values were not imputed, and data were analyzed according to an available-case principle. All tests were two-tailed and p-values < 0.05 were considered statistically significant.

3.2.5.1 Measurement internal consistency

In order to assess the internal consistency of HADS, EQ-5D, and ODI questionnaires for PSPS-T2 population, Cronbach's alpha and its confidence interval were calculated for each measurement questionnaire. Internal consistency measures whether several items from the same questionnaire measure the same general construct based on the amount of shared information. Internal validity of these questionnaires have never been studied for the PSPS-T2 population. Confidence intervals were estimated using the Bootstrap technique [121].

3.2.5.2 Correlation between EQ-5D, NPRS, ODI and HADS depression scores

In order to estimate the correlations between our measurement variables, Pearson's simple correlation was calculated between the different variable assessments at baseline and 12-month follow-up. Baseline and 12-month follow-up correlations were compared using a z-test after conducting Fisher transformation on the coefficients.

3.2.5.3 Clustering of the impact of pain intensity, functional disability, and depression on HRQoL

A mixture of mixed effect models [80] was used in order to extract the latent classes based on the effects of functional disability (ODI score), pain intensity (NPRS) and depression (HADS depression score) on HRQoL (EQ-5D index). The intercept was fixed between the classes. In the mixed effects model of the mixture, the explained variable was HRQoL measured by the EQ-5D index at baseline, 3, 6, 9 and 12-month follow-up. The explanatory variables were global NPRS, ODI score and HADS depression score also measured at baseline, 3, 6, 9 and 12-month follow-up. We included a random intercept effect to account for intra-patient correlation. The model could be written as:

$$EQ-5D_{ij} = \beta_0 + \mu_{0i} \sum_{c=1}^{C} \mathbb{1}_{\{z_i=c\}} (\beta_{1c} ODI_{ij} + \beta_{2c} NPRS_{ij} + \beta_{2c} HADS_{ij}) \qquad (3.1)$$

where i and j are the indexes associated with the patient and follow-up visit respectively. The optimal number of latent classes was identified using the Bayesian Information Criterion (BIC) [122]. We estimated the model for different numbers of classes (ranging from 1 to 4 classes) and the model with the smallest BIC was considered as the final one. Standardized coefficients and their 95% confidence intervals were reported. An implementation of the mixture of mixed effect models is available in the R package lcmm [84].

3.2.5.4 Baseline characteristics influencing class membership

After the classes and their respected effects were estimated, they were used as a binary dependent variable (the optimal number of extracted classes was 2) to determine the impact of sociodemographic, psychological and cognitive-behavioral variables on class membership. Normality of distributions was verified using the Shapiro-Wilk test. Relationship between qualitative variables and class membership was verified using the Fisher exact test for non-ordinal qualitative variables and the Cochran-Armitage test for ordinal variables. The Wilcoxon rank-sum test was used to test the relationship between class membership and quantitative variables. NPRS, ODI percentage, HAD depression and anxiety scales, gender, level of education, CSQ catastrophizing, ignorance (ignoring pain sensation), reinterpretation (reinterpreting pain sensations), diversion (diverting attention from pain) and self-encouragement subscales, perceived arduous working conditions, and the interaction between gender and perceived arduous working conditions at baseline were included in a logistic regression model in order to assess the adjusted effects of these variables on class membership. Model coefficients and their 95% confidence intervals were reported.
We also studied the relationship between level of education, perceived arduous working conditions and coping strategy scores (Pearson correlation coefficients), in order to help us to interpret our bivariate and multiple regression results.

3.3 Results

3.3.1 Study population

Out of the 200 patients enrolled, 168 (84%) completed data at 3-month, 166 (83%) at 6-month, 145 (72.5%) at 9-month and 146 (73%) at 12-month follow-up. Two patients completed only the sociodemographic and clinical data, but did not complete any of the evaluation questionnaires. Therefore, they were used only for descriptive analysis and were removed from the other analyses. The relatively high percentage of dropout was due to the fact that the study was observational and the treatment the patients receive does not change whether they participated or not in the study.

3.3.2 Baseline characteristics

Table 3.1 shows the demographics of the study population. The study participants mean age was 53 (13) years old and 111 (56.1%) were females. Ninty-nine patients (50.0%) had at least two surgeries. At baseline, the study sample had a mean EQ-5D of 0.27 (0.24), global pain

NPRS of 6.1 (1.5), a mean ODI percentage of 34.3% (11.7) and a mean HAD depression score of 8.6 (3.9).

Variables	n	%
Age mean (sd)	53(13)	
Gender		
Male	87	43.9
Female	111	56.5
Professional status		
In professional activity	41	20.7
Retired	38	19.2
Disability	38	19.2
Sick leave	38	19.2
Long-term illness	16	8.1
Unemployment	7	3.5
Without professional activity	20	10.1
Educational level		
≤ 12 years	153/193	79
> 12 years	40/193	21
Number of spinal surgeries		
1	99	50.5
2	59	29.5
3	28	14
4	8	4
5+	4	2
Neuropathic pain (DN4* score ≥ 4)		
Yes	147/184	80
No	37/184	20
Pain duration		
≤ 5 years	46/198	23
> 5 years	152/198	77

TABLE 3.1 – Patients baseline sociodemograhic and clinical characteristics ($n = 198$).
* DN4: Douleur Neuropathique en 4 - questions, a 10-item questionnaire with a score ranging from 0 to 10, used to diagnose neuropathic pain.

3.3.3 Measurements correlation

The assessment tools used in this study have shown good internal consistency. At baseline, EQ-5D had a Cronbach alpha of 0.693 ($CI95\% = [0.605, 0.758]$), the ODI questionnaire had a Cronbach alpha of 0.801 ($CI95\% = [0.743, 0.848]$) and the HADS depression questionnaire had a Cronbach alpha of 0.775 ($CI95\% = [0.716, 0.814]$).
The correlation coefficients between the different measurements for baseline and 12-month follow-up are shown in Table 3.2. At baseline, there was a high negative correlation

between EQ-5D quality of life index and ODI disability score. EQ-5D index was also moderately correlated with HADS depression score, pain intensity NPRS and CSQ catastrophizing score. Disability was also positively correlated with depression score, pain intensity and catastrophizing score. Pain intensity was moderately correlated with depression score and weakly correlated with catastrophizing score. Correlations between EQ-5D index and the different measurements increased between baseline and 12-month follow-up. The increase between baseline and 12-month follow-up in correlation with EQ-5D index was significant for NPRS (p = 0.011) and ODI score (p = 0.019).

Baseline

Variables	ODI	NPRS	HADS dep^T	CSQ cata$^+$
EQ-5D	-0.66^{***}	-0.35^{***}	-0.44^{***}	-0.44^{***}
ODI		0.44^{***}	0.40^{***}	0.33^{***}
NPRS			0.33^{***}	0.26^{**}
HADS dep^T				0.50^{***}

1-year follow-up

Variables	ODI	NPRS	HADS dep^T	CSQ cata$^+$
EQ-5D	-0.77^{***}	-0.55^{***}	-0.56^{***}	-0.40^{***}
ODI		0.54^{***}	0.57^{***}	0.42^{***}
NPRS			0.39^{***}	0.21
HADS dep^T				0.61^{***}

^{T}HADS depression subscale. $^+$CSQ catastrophizing subscale.
$^*p < 0.01$, $^{**}p < 0.001$, $^{***}p < 0.0001$.

TABLE 3.2 – Correlations between EQ-5D index, ODI, HADS depression score and CSQ catastrophizing at baseline and at 12-month follow-up.

3.3.4 Standard 1-class mixed effects model results

Our one component model showed that disability, depression and pain intensity had a statistically significant effect on HRQoL, the effect of disability being the strongest according to the estimated standardized coefficients. The one-component mixed-effect model can be found in Table 3.3. The estimated adjusted standardized effect of disability on HRQoL was $\beta_{ODI} = -0.48$ ($p < 0.0001$). The effect of pain intensity on HRQoL was $\beta_{NPRS} = -0.13$ ($p < 0.0001$). The effect of depression on HRQoL was $\beta_{HADS} = -0.20$ ($p < 0.0001$).

3.3.5 Two-class mixed model results

Two latent classes were extracted from the mixture of mixed effects models. The two classes were well separated as the posterior mean probability of class 1 was 0.754 and the mean posterior probability of class 2 was 0.763. Table 3.3 shows the results of the standard mixed effects model and the mixture of mixed effects models in detail. Mean trajectories for class 1 and class 2 patients are represented in figure 3.1.
The standardized coefficients, their confidence intervals and their significance level of the

two-component mixed-effect model can be found in Table 3.3. For class 1 patients, consisting of 68.7% of patients (136/198), the effect of pain intensity on HRQoL was not statistically significant (Std beta = 0.039, $p = 0.46$). The effects of disability and depression were statistically significant (Std beta = -0.76, $p < 0.0001$ for disability and Std beta = -0.19, $p < 0.0001$ for depression).
For class 2 patients, consisting of 31.3% of study participants (62/198), both pain intensity and depression had a significant effect on HRQoL (Std beta = -0.35, $p < 0.0001$ for pain intensity and Std beta = -0.22, $p = 0.0001$ for depression). However, ODI percentage did not have a statistically significant effect (Std beta = -0.11, $p = 0.23$).
From now on, we will refer to class 1 as "disability class" and class 2 as "pain intensity class".

Global model			
Variables	Standardized coefficient	Standard error	p-value
Intercept	-0.0029	0.037	0.93
ODI percentage	-0.48	0.034	< 0.0001
NPRS	-0.13	0.027	< 0.0001
HADS depression	-0.20	0.030	< 0.0001
Class 1 model			
Variables	Standardized coefficient	Standard error	p-value
Intercept	-0.0060	0.037	0.94
ODI percentage	-0.76	0.074	< 0.0001
NPRS	0.039	0.041	0.46
HADS depression	-0.19	0.044	< 0.0001
Class 2 model			
Variables	Standardized coefficient	Standard error	p-value
Intercept			
ODI percentage	-0.11	0.095	0.23
NPRS	-0.35	0.065	< 0.0001
HADS depression	-0.22	0.057	0.0001

TABLE 3.3 – Global, class1 and class2 mixed effects models representing the effects, functional disability, pain intensity and depression have on health related quality of life

3.3.6 Factors influencing the class membership

Bivariate and logistic regression analysis of the relationship between class membership and sociodemographic and behavioral variables are presented in table 3.3. The bivariate analysis revealed that higher educational level was associated with higher probability of belonging to “disability class” ($p = 0.004$). Similarly, The mean catastrophizing CSQ scores was higher in the "disability class" than in the "pain intensity class" but it was only significant at a 10% level ($p = 0.09$). Sex, age, pain duration, ODI percentage, perceived arduousness of work had no statistically significant impact on class membership in the bivariate analysis ($p > 0.22$). In the multivariate logistic regression analysis, patients with a higher educational level quality of life is more affected by disability than pain intensity ($p = 0.003$). Patients with higher catastrophizing scores were also more impacted by disability more than pain intensity ($p = 0.021$). The interaction between sex and physical labor was significant at a 10% level. Males that perceive their work as physical where more impacted by disability than by pain intensity ($p = 0.080$ for the interaction term) which was not the case of females working in physical labor, which were significantly more impacted by pain intensity ($p = 0.031$).
In order to help interpret our results, we also studied the relationships between the different factors studied in the multivariate and bivariate analysis. We found that the level of study did not have a significant relationship with the perceived arduousness of work ($p = 0.71$). There is a significant relationship between CSQ catastrophizing score and the perceived arduousness of work ($p = 0.029$). Non-working patients had the highest CSQ catastrophizing score followed by working patients who perceive their job as physical and finally working patients who do not perceive their job as physical. There is a low negative non-significant correlation between catastrophizing and educational level. Patients with lower catastrophizing scores had a higher educational level ($r = -0.13$, CI95% $= [-0.28, 0.02]$, $p = 0.09$).

Variable	Mean (sd)/n(%)		Standardized coefficients	95% Confidence interval	Adjusted p-value
	Class 1 $n = 136$	Class 2 $n = 62$			
Intercept			-0.969	$[-1.353, -0.584]$	< 0.001
global NPRS at baseline	5.91 (1.41)	6.47 (1.57)	0.428	$[-0.023, 0.878]$	0.055
ODI percentage at baseline	33.92 (10.83)	34.98 (13.40)	0.051	$[-0.418, 0.519]$	0.83
Sex (male)	60/136 (44%)	27/62 (44%)	0.057	$[-0.376, 0.490]$	0.79
Age (years)	52.01 (12.08)	54.47 (13.41)	0.057	$[-0.369, 0.484]$	0.79
Current pain duration (years)	4.43 (5.92)	4.64 (6.35)	0.115	$[-0.261, 0.490]$	0.54
Perceived physical job					
working in a job perceived as physical	17 (13%)	9 (15%)	0.550	$[0.036, 1.064]$	0.031
working in a job not perceived as physical	10 (7%)	5 (8%)	0.132	$[-0.331, 0.595]$	0.57
not in a professional activity	109 (80%)	48 (77%)	-	-	-
Level of study (years)	11.36 (3.14)	9.47 (4.36)	-0.602	$[-1.005, -0.199]$	0.003
CSQ catastrophizing score	14.17 (4.64)	12.98 (4.18)	-0.509	$[-0.952, -0.065]$	0.021
Sex (male)*physical job					
Male working in a physical job	12/60 (20%)	2/27 (7%)	-0.520	$[-1.119, 0.079]$	0.080
Male working but not in a physical job	3/60 (5%)	2/27 (7%)	-0.100	$[-0.540, 0.340]$	0.65
Male and is not in a professional activity	45/60 (75%)	23/27 (86%)	-	-	-

TABLE 3.4 – Demographic and cognitivo-behavioral factors associated with class membership

3.4 Discussion

Applying a finite mixture mixed effects regression model to this multicentric observational prospective study allowed us to cluster PSPS-T2 population into two distinct classes depending on the effects of pain intensity, disability, and psychological distress on HRQoL. The first class, named "disability class", represented patients with HRQoL primarily impacted by functional disability and depression, while the HRQoL of the second class, named "pain intensity class", was more impacted by pain intensity and depression. Furthermore, level of education, perceived arduous working conditions, anxiety/depression, and coping strategies were significantly involved in the characterization of disability and pain intensity classes.

3.4.1 A need for multidimentional composite pain assessment to represent HRQoL heterogeneity

Our results showed that psychological distress was significantly involved in the HRQoL of all the PSPS-T2 patients enrolled in this study. In addition to psychological distress, HRQoL was significantly impacted by functional disability for 68.7% of patients, and by pain intensity for 31.3% of the PSPS-T2 patients. All in all, our findings suggest that there are significant differences in the importance of sensory, functional, and emotional components in PSPS-T2 patients, which should be considered to propose specific care pathways according to a given patient and his or her intrinsic pain characteristics. These results indicate that chronic pain evaluation should not be exclusively evaluated based on pain intensity.
Corroborating these findings, the multidimensionality of chronic pain is well-established in the literature [19, 23, 30] and reinforced by the bio-psycho-social model [123, 124]. However, pain management remains mainly focused on pain intensity relief. The vast majority of comparative pain research works are based on primary outcomes focused on pain reduction scores (Visual Analogic Scale (VAS) or NPRS). Harmful consequences of focusing exclusively on pain intensity have been observed by Ballantyne and Sullivan [19], who contended that the multimodal pain management approaches proposed to chronic pain patients should include behavioral and physical-rehabilitation, which cannot be adequately evaluated using only pain intensity measures [125]. We have shown that not only pain is multi-factorial but also that patients responded differently to changes in pain intensity and functional disability depending on their sociodemographic and behavioural characteristics. In a review [23], the authors claimed that it is not necessary to reduce pain intensity systematically in order to achieve adequate chronic pain management. This suggestion was supported by several studies indicating that multidisciplinary rehabilitation programs have shown greater effects on disability and quality of life than on pain intensity [126, 127]. Further considerations could also include that patients should be grouped into profiles in order to obtain better estimates of the impact of the different pain dimensions on the patient.

3.4.2 Characterisation of HRQoL classes

The literature has shown that cultural and sociodemographic characteristics are related to how patients perceive their pain and to how much it impacts on their quality of life

[128–130]. In a qualitative literature review including 77 articles on patients with general chronic pain, Samulowitz et al. [130] synthesized the gender biases often present in a patient-clinician relationship. The authors found that for men presenting with chronic pain, work and being a "breadwinner" were important and were linked to their sense of masculinity. The authors also found that physical activity, as an important part of male identity, was a recurring theme when men were describing their experience of living with pain. These findings confirm the hypothesis that functional capacity plays a significant role in male HRQoL patients, especially those with a physical job. In our study, we quantitatively found that the HRQoL of male patients who perceived their job as "physically demanding" was more impacted by functional disability than by pain intensity. For these patients, a deterioration in functional capacity generates a freeze in their professional activity and heavily impacts both self-image of "breadwinner" and their financial status [129]. Although physical labor is more often associated with lower educational level, we found, in contrast, that the HRQoL of patients with lower educational level was more impacted by pain intensity than functional disability. In this study, we did not find a significant relationship between educational level and how patients perceive their work (physical/not physical) (p=0.71). Roth and Geisser [128] have shown, in a study including 299 patients with chronic spinal pain, that patients with low educational levels had greater belief that pain is disabling and uncontrollable. Patients with high educational level were more likely to participate in activities and to adopt more adaptive coping strategies [128]. Patients with higher education were more physically active and considered sports and physical activity as an active coping strategy. This indicates that for patients with higher educational level, functional capacity preservation is necessary in order to cope with pain, which might explain why their HRQoL is more impacted by functional disability than by pain intensity. Similarly, we found that the HRQoL of patients who frequently used adaptive coping strategies such as diverting attention from pain or reinterpreting pain were more affected by functional disability than by pain intensity. These results indicate that patients who use either emotion-based or active adaptive coping strategies are more impacted by functional disability. In a study of 103 patients with chronic PSPS-T2 pain [131], the authors found that physical activity practice was negatively associated with the use of maladaptive coping strategies such as catastrophizing. These findings suggest that a decrease of functional capacity might lead patients not to be able to cope with pain in a healthy manner, finally leading to deterioration in HRQoL. These results indicate that gender, work's physical burden, level of education, and coping strategies are important factors that need to be considered during HRQoL assessment and in the process of identifying what is important for the patient.

3.4.3 Study strengths and limitations

Aside from the relatively large sample size and the prospective design, the strengths of this study also included the novel application of model based statistical clustering. This allows overcoming the use of subjective goal identification tools. When patients are asked about what impacts them the most, the majority of patients tend to aim for pain intensity relief, as they believe that functional improvement is a necessary following of pain relief [132]. In our study, we found that the majority of patients' quality of life is affected by disability

more than pain perception. Statistical techniques allow for an identification of profiles based on the overall behavior and not the individual behavior but they allow for an objective patient-profile identification, which is not achievable using self-reported tools and questionnaires.
Despite the study strengths, some limitations must be considered. First, our study sample was exclusively drawn from a PSPS-T2 population, which did not allow us to extend our findings to the general pain population, including acute and chronic general or widespread pain and non-operated patients. Another limitation is that we focused solely on pain intensity, functional disability and depression when describing HRQoL. Although it is clear that these three components represent major dimensions of quality of life, having significant consequences on patient HRQoL, there are other chronic pain components, which we deliberately did not include in our study: expectations, social insecurity or lack of sleep, and which might also contribute to building on a holistic HRQoL model. A new cohort study including patients with widespread acute and chronic pain, aiming to focus with more clarity on all dimensions of quality of life could yield more specific biopsychosocial clustering and optimized characterization of pain-related quality of life. Furthermore, we assume that the extension of this pilot work should be based on a larger cohort, as a prospective registry, to bridge this gap and address two specific goals: external validation and expansion of the results of this study, since there is an obvious lack in literature regarding how chronic pain patients belonging to distinct latent subpopulations are affected individually by different components of pain.

3.4.4 Therapeutical implications

From a clinical perspective, the identification of patient clusters, determined from their social and cognitive-behavioral characteristics, could help physicians to tailor personalized care, to propose "à la carte" programs and ultimately reinforce therapeutic alliance, thereby improving patient outcomes. Patients highly impacted by their functional disability could be preferentially oriented towards therapies and programs that have demonstrated a specific impact on disability such as cognitive-behavioral-based physical therapies [133, 134], to cope with their body, while patients predominantly impacted by pain intensity could be oriented towards specific pain relief therapies, to cope with pain perception, emotions and revisit their mind. It appears clearly, from the extreme variability of treatments available to patients presenting with chronic pain, that each treatment strategy has a preferential impact on pain intensity or on functional disability [101]. Given these findings, we should probably insist on patient profiling to redesign personalized pain patient care pathways and to put the patient back at the center of the pain puzzle.

3.5 Conclusion

Applying a mixture of mixed effects models to data of chronic pain patients might lead to the development of new pain evaluation strategies and to a better understanding of HRQoL dimensions. The results of this study can be considered as a starting point to refine a multidimensional, personalized, cluster-based, composite evaluation of HRQoL in chronic pain patients, which could impact on the optimization of pain management.

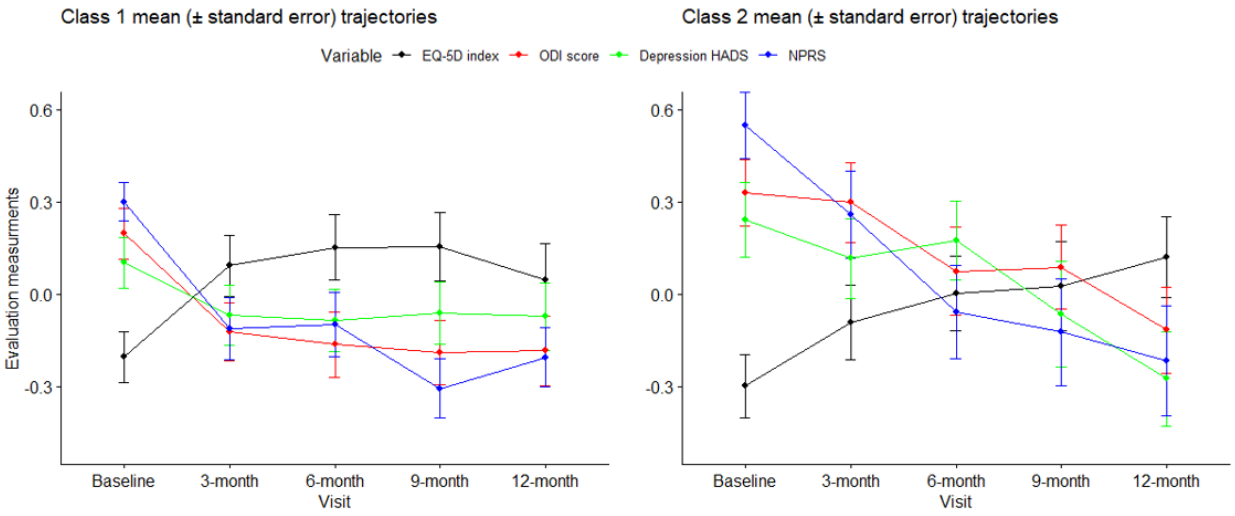

FIGURE 3.1 – Class 1 and class 2 mean trajectories for each evaluation criterion at baseline and at 3-month, 6-month, 9-month and 12-month follow-ups. EQ-5D: EuroQol-5 Dimensions, ODI: Oswestry Disability Index, NPRS: Numeric Pain Rating Scale, HADS: Hospital Anxiety and Depression Scale. We can observe that for class 1 (disability class) patients, the increase in quality of life (black line) is more associated with a decrease in functional disability (red line). However, for class 2 (pain intensity class) patients, we can observe that an increase in quality of life is more associated with a decrease in pain intensity.

Chapitre 4

Mixture of varying-coefficient models with random effects processes for intra and inter-subject effect estimation

Résumé du chapitre en français

Des quantités massives de données longitudinales sont devenues disponibles grâce aux études de cohorte de grandes échelles. Les techniques de modélisation paramétriques et non-paramétriques dédiées aux données longitudinales sont donc très utiles. Les modèles à coefficients variables ont récemment attiré beaucoup d'attention grâce à leur capacité à modéliser les liens potentiels entre une variable de réponse variant dans le temps et des covariables variant dans le temps. Dans les modèles à coefficients variables, les coefficients de régression sont représentés par des fonctions lisses du temps. Cependant, les modèles à coefficients variables sont généralement utilisés sans tenir compte de l'hétérogénéité des données concernant la potentielle existence de plusieurs sous-échantillons non observés qui sont générés selon des modèles différents. Modéliser de telles données en supposant que l'échantillon provient d'une population homogène correspond à un biais majeur due à une mauvaise spécification du modèle (*i.e.* la non-inclusion des classes latentes représentant les sous-échantillons homogènes). Par exemple, résumer un échantillon issue d'un mélange de lois gaussiennes par une simple moyenne empirique peut aboutir à de mauvaises interprétations. Pour combler cette limitation, nous proposons dans ce chapitre un mélange de modèles à coefficients variables avec des effets aléatoires modélisés comme des processus stochastiques. Nous avons développé une procédure *backfitting* pour estimer notre modèle. Une méthode de validation croisée a également été utilisée pour la sélection des hyperparamètres de lissage, qui influencent la performance du modèle. Le modèle proposé a été évalué sur des données simulées et réelles. Le modèle proposé dans ce chapitre est une extension non-paramétrique du modèle proposé par [80] appliqué dans le chapitre 3. Concernant les données réelles, le modèle a été appliqué sur la base de données

PREDIBACK, afin d'estimer l'évolution dans le temps de l'effet de l'intensité de la douleur, de la capacité fonctionelle et de la dépression sur la qualité de vie des patients douloureux chroniques. Cela nous a permis de contraster les résultats de notre modèle avec ceux obtenus avec un mélange de modèle à effets mixtes invariants dans le temps, obtenus dans le chapitre 3. Les données simulées et réelles nous ont permis de montrer que le modèle proposé dans ce chapitre clusterise les sujets en groupes homogènes et permet d'obtenir de meilleures estimations des effets intra et inter-sujets par rapport aux autres approches développées dans la littérature.

4.1 Introduction

Longitudinal Data enables new perspectives to emerge, particularly in the field of medicine. It has become available thanks to the development of efficient data collection methods, facilitating the emergence of large-scale, long-term cohort studies [1]. Thus, statistical methods applied to such data appear of great interest.

Researchers can model longitudinal data in several ways including parametric models such as mixed effect models [2, 42] and non-parametric models such as varying-coefficient models [39].

In the classical framework of mixed effect models, fixed effects do not vary over time, unless the model includes a time predefined function as a covariate. On the other hand, varying-coefficient models estimate these effects using unknown smooth time functions estimated by non-parametric methods. Each of these two algorithms has its advantages and disadvantages. Mixed effect models incorporate intra-subject covariances. This is not the case with varying-coefficient models, since non-parametric models are estimated locally, which complexifies the incorporation of the covariance structure [135–137]. However, varying-coefficient models generate smooth functions representing effect evolution over time, which gives more information than the single constant effect estimates provided by mixed effect models.

Several authors have proposed extensions of time-varying coefficient models for longitudinal data [53, 62, 70, 136]. Hoover *et al.* [53] and Wu *et al.* [62] proposed various estimation procedures for varying-coefficient models for longitudinal data but did not incorporate the information on the covariance structure within subjects directly into their models. Sun *et al.* [136] proposed a time-varying coefficient model where they included time-constant parametric random effects. Wu and Liang [70] proposed a varying-coefficient model where they included time-varying random effects represented by random processes.

In their classical form, both mixed effects and varying-coefficient models are unable to model heterogeneous data gathered from different unknown subpopulations. For heterogeneous data, it would be more appropriate to estimate different effects for each group of subjects depending on the subpopulation they were drawn from. Mixture of regression models are generally used to solve this problem [40]. Mixture of regressions generates a different model for each homogeneous subpopulation. Since these subpopulations are unknown, mixture models are generally estimated using the Expectation Maximization (EM) algorithm.

In this chapter, we propose a new model that incorporates a time-varying mixture of inter-subject effects and intra-subject effects. We achieved this by using a mixture of varying-coefficient models with random effects, where each subpopulation is modeled by distinct smooth functional population-effects and individual intra-subject effects represented by random functions. In section 4.2, we describe our proposed model and its parameters. In section 4.3 and section 4.4, we describe the modified EM algorithm used to estimate our model. In addition, in section 4.6, we propose a Cross-Validation (CV) procedure used to estimate the optimal smoothing parameters of the model. We also propose and test a modified BIC criterion used to estimate the number of mixture components in section 4.5. We present the results of both simulated and real data examples in section 4.7. Using simulated data, we compared our model to the mixture of

varying-coefficient model with no random effects proposed by *Huang et al.* [85]. In our real data examples, we used the PREDIBACK study dataset to estimate the time-varying effects of pain perception, disability and depression on chronic pain after spinal surgery patients (section 4.7.2).

4.2 Mixture of varying-coefficient models with random effects (MVCRE)

In this section, we describe the statistical motivation behind our model and give a detailed description of its mathematical formulation and its parameters.

Throughout this chapter, we suppose that we have N_s subjects. For each subject i, we observe data at N_i time points, making a total of $N = \sum_{i=1}^{N_s} N_i$ observations. At the j^{th} time point, we have a measure of $Y_i(t_{ij})$, which is a continuous outcome of interest for subject i at time t_{ij}, and we have a vector $X_i(t_{ij})$ of length $p+1$. The first element of $X_i(t_{ij})$ is equal to 1 and the the other elements are the values of the p explanatory variables $(X_{ki}(t_{ij}))_{k=1,\dots,p}$ at time t_{ij} (*i.e.* $X_i^T(t_{ij}) = (1, X_{1i}(t_{ij}), \dots, X_{pi}(t_{ij}))$). The scalar 1 allows the estimation of the time varying intercepts. Our aim is to model the effect of the explanatory input variables $(X_k(t))_{k=1,\dots,p}$ on the outcome of interest $Y(t)$. Time-varying coefficient models can be used for this purpose but these models have two fundamental limitations. The first limitation is that they do not take into account the within-subject correlation that is generally present in longitudinal data. In longitudinal studies each subject has some intrinsic characteristics that produce alterations from the general/population effects which are not taken into account by the error terms. The non-inclusion of these within-subject effects in the modeling process produces biased estimates of the population effects due to an underspecified covariance structure [138]. In order to address this issue, Wu and Liang [70] proposed a varying-coefficient model with random effects for which intra-subject effects are estimated using time-varying random effects represented by random functions. Their model can be written as:

$$Y_i(t_{ij}) = X_i^T(t_{ij})\beta_i(t_{ij}) + e_i(t_{ij}),$$

where $\beta_i(t) = \beta(t) + \mu_i(t)$, $\beta(t)$ are smooth functions representing the time-varying fixed effects, $(\mu_i(t))_{i=1,\dots,N_s}$ are independent random effect processes with mean 0 and covariance function $\rho(t,s) = \text{Cov}(\mu_i(t),\, \mu_i(s))$ for $(t,s) \in \mathbb{R}_+^2$, and $(e_i(t))_{t\in\mathbb{R}_+}$ is a stochastic process with a constant mean 0 and a variance function $\sigma^2(t)$. We make the assumption that the error processes $(e_i(t))_{i=1,\dots,N_s}$ are independent of the random effect processes $(\mu_i(t))_{i=1,\dots,N_s}$, similarly to the classical mixed effects models.

The second limitation of the varying-coefficient models is that one global time-varying coefficient is estimated for each time-varying explanatory variable even when the data shows evidence of groups of subjects who differ in the effects of the explanatory variables on the outcome of interest. We refer to these differences as inter-subject heterogeneity. This problem is generally addressed using mixture models. These models are a clustering approach that estimates different effects for different latent (not observed) subgroups. This is achieved by incorporating a latent class variable that represents the subgroups to which each subject belongs. In order to incorporate mixture modeling into varying-coefficient

models, Huang *et al.* [85] proposed a time-varying mixture of varying-coefficient models that have the following mathematical formulation:

$$Y_i(t_{ij}) = \sum_{c=1}^{C} \mathbb{1}_{\{z_i(t_{ij})=c\}} \left(X_i^T(t_{ij})\beta_c(t_{ij}) + e_{ci}(t_{ij}) \right),$$

where $z_i(t)$ is a categorical latent process representing the latent class of subject i at time t. At a time t, $z(t)$ takes values in $\{1, \ldots, C\}$ associated with the probabilities $\{\pi_1(t), \ldots, \pi_C(t)\}$ ($\sum_{c=1}^{C} \pi_c(t) = 1$ for any t). The indicator function $\mathbb{1}_{\{z_i(t)=c\}}$ is equal to 1 if subject i belongs to class c at time t and 0 otherwise. The smooth functions $\beta_c(t)$ represent time-varying fixed effects for subjects in the class c at time t. The error term for class c, $e_{ci}(t)$ is a stochastic process, with mean 0 and variance function $\sigma_c^2(t)$.
We propose in this chapter a natural extension to the varying-coefficient model with random effects and the mixture of varying-coefficient models. We assume a mixture of varying-coefficient models with time-varying random effects. Hence we propose in this chapter the model given by:

$$Y_i(t_{ij}) = \sum_{c=1}^{C} \mathbb{1}_{\{z_i(t_{ij})=c\}} (X_i^T(t_{ij})\beta_{ci}(t_{ij}) + e_{ci}(t_{ij})), \tag{4.1}$$

where $\beta_{ci}(t) = \beta_c(t) + \mu_i(t)$, $\beta_c(t)$ are smooth functions representing time-varying fixed effects for subjects in the class c at time t. Similarly to the varying-coefficient model with random effects, the random functions $(\mu_i(t))_{i=1,\ldots,N_s}$ have a mean 0 and covariance function $\rho(t,s) = \text{cov}(\mu_i(t), \mu_i(s))$. These random functions $\mu_i(t)$ model the deviation from the sample mean effects $((\beta_c(t))_{c=1,\ldots,C})$ for subject i.
In model (4.1), the parameters to be estimated are the time-varying fixed effects $(\beta_c(t))_{c=1,\ldots,C}$, the variance components $(\sigma_c(t))_{c=1,\ldots,C}$, the random effects $(\mu_i(t)))_{i=1,\ldots,N_s}$ and finally, the mixture proportions $(\pi_c(t))_{c=1,\ldots,C}$, which are the proportions of the latent classes at a time point t.
There are two differences in our model compared to the classical time varying-coefficient models. The first difference is the inclusion of time-varying random effects given by the random functions $\mu_i(t)$. The random processes $\mu_i(t)$ are the deviations from the inter-subject effects $\beta_c(t)$ for subject i at time t. just like in the classical mixed effects models, adding random effects can lead to smaller standard errors due to the phenomenon known as shrinkage [139] that we observe later in our simulations. It also allows a wider range of research to be investigated including the estimation of within level-effects. In addition, from an applicative point of view, real data from the same subjects generally carry some intra-subject correlation. This means that the covariance structure of a model with only time-varying fixed effects would be underspecified in case of such data. For example, in medical research with subjective evaluation tools such as patient reported outcomes, the relationship between variables might depend heavily on properties that are intrinsic to the patient.
In order to examine and model inter-subject heterogeneity, we included time-varying mixture of latent classes represented by the indicator functions $\mathbb{1}_{\{z_i(t)=c\}}$ for c in $1, \ldots, C$. The time-varying fixed effects are specific to each latent class but are different between classes at each time point.

In the next section, we will show how these parameters can be estimated using a two-step iterative procedure where the proposed model is separated into two components that are estimated iteratively.

4.3 Parameters Estimation Procedure

The most straightforward procedure to estimate the model (4.1) parameters is to estimate locally, for a grid of time points, a mixture of mixed effect models given by:

$$Y_{ij} = \sum_{c=1}^{C} \mathbb{1}_{\{z_{ij}=c\}}(X_{ij}^T \beta_{ci} + e_{cij}),$$

where Y_{ij} denotes $Y_i(t_{ij})$ and X_{ij} denotes $X_i(t_{ij})$.
Smoothing can then be applied to obtain the regression functions, the time-varying mixture proportions and the variance/covariance functions. The problem with this procedure is that it leads to the label switch problem [140], which arises because class membership proportions are time-varying. More precisely, it occurs when the class labels switch from a time point to another which means that the model does not estimate the real coefficients but rather the permutations (in time) of these coefficients. The conventional EM algorithm estimates the non-parametric functions separately for a set of grid points, making it difficult to assign the same class labels for these estimators across all grid points. Huang *et al.* [85] proposed a modified EM algorithm to estimate a varying mixture of varying-coefficient (MVC) models while avoiding the problem of label switch. The method Huang et al proposed is to estimate globally the probability of membership to a class for all data points.
The modified EM they proposed is used to estimate the parameters of a mixture of varying-coefficient model with no random effects: $Y_i(t_i) = \sum_{c=1}^{C} \mathbb{1}_{z_i(t_i)=c}(X_i^T(t_i)\beta_c(t_i) + e_{ci}(t_i))$. Contrary to the model we proposed, this model does not consider intra-subject dependence (all observations are considered independent). The modified EM algorithm proposed by Huang et al can be found in algorithm 1.
The quantity (4.2) in algorithm 1 is the weighted conditional expectation of the complete log-likelihood of a finite mixture model with the parameters β and σ (conditional on the observed data). The maximization of (4.2) gives us the local estimates $\hat{\beta}_c$ and $\hat{\sigma}_c$ that we interpolate to get our estimates of $\beta_c(t)$ and $\sigma_c(t)$.
Kernel functions K_h are used in non-parametric local regression as a measure of similarity such that similar points are weighted higher when constructing the regression functions/coefficients locally. The bandwidth h helps set the amount of closer points that are given weights. Low bandwidth means that only close points contribute to the local estimation which might result in overfitting. Higher bandwidths allow distant points to be given non-null weights which might result in oversmoothed coefficients. We propose later in this chapter a method to identify the optimal global bandwidth using CV.
To estimate the parameters of the model introduced in this chapter, we propose an approach that separates the model parameters estimation into two distinct estimation methods by iterating between the estimation of random effect processes and the estimation

Algorithm 1 EM algorithm for estimating a time-varying mixture of varying-coefficient models without random effects

Iterate the following two steps until convergence of the algorithm:
E-step: For each observation $i = 1,\ldots,N$ and each component $c = 1,\ldots,C$, calculate:

$$r_{ci} = \frac{\pi_c(t_i)\phi(Y_i|\mathrm{X}_i^T\beta_c(t_i),\sigma_c(t_i))}{\sum_{k=1}^{C}\pi_k(t_i)\phi(Y_i|\mathrm{X}_i^T\beta_k(t_i),\sigma_k(t_i))},$$

where $\phi(Y|\mathrm{X}^T\beta_c(t_i),\sigma_c(t_i))$ is the conditional density function of Y, given $\beta_c(t_i)$ and $\sigma_c(t_i)$ at the point Y_i. r_{ic} is the probability that Y_i was generated by the component c given the models' likelihood, also called class membership probabilities. The quantity $\pi_c(t_i)$ is the probability of an observation to belong to the component c, at time t_i.

M-step: For $c = 1,\ldots,C$, and t in a set of grid points, calculate:

$$\hat{\pi}_c(t) = \frac{\sum_{i=1}^{N} r_{ci}K_h(t_i - t)}{\sum_{i=1}^{N} K_h(t_i - t)},$$

where $\hat{\pi}_c(t)$ is the Nadaraya-Watson estimator [51, 52] of $\pi_c(t)$ associated with kernel $K_h(t_i - t) = K(\frac{t_i - t}{h})$, where K is a kernel function. A kernel is a positive real-valued integrable function that has an integral of 1 over $\mathbb{R}$ and is symmetric (*i.e.* $K_h(t) = K_h(-t)$).
We then update $\beta_c(t)$ and $\sigma_c(t)$ by maximizing:

$$\sum_{i=1}^{N}\sum_{c=1}^{C} r_{ci}\log[\phi(Y_i|\mathrm{X}_i^T\beta_c,\sigma_c)]K_h(t_i - t). \tag{4.2}$$

of time-varying fixed effects similarly to the backfitting method that was introduced for estimating additive models [141]. In our algorithm, at each iteration, we replace the density function $\phi(Y_i|\mathrm{X}_i^T\beta_c(t_i),\sigma_c(t_i))$ by $\phi(Y_i|\mathrm{X}_i^T\beta_c(t_i) + \mathrm{X}_i^T\hat{\mu}_i(t_i),\sigma_c(t_i))$, where $\hat{\mu}_i(t_i)$ are the estimated random effects from the previous iteration. The random effects are estimated locally using a mixed effects model.
The estimation procedure of the parameters $\beta_c(t)$, $\pi_c(t)$ and $\sigma_c(t)$ for $c = 1,\ldots,C$ and $\mu_i(t)$ for $i = 1,\ldots,N_s$ is summarized in algorithm 2.

In step 4 of the algorithm, by "best fit" we mean the model leading to a maximized log-likelihood (*i.e.* $\underset{c\in\{1,\ldots,C\}}{\arg\max}\,\phi(\overline{Y}_i(t_{ij})|\mathrm{X}_i^T(t_{ij})\beta_c(t_{ij}),\sigma_c(t_{ij}))$).

In step 5, by setting $\mu_i(t) = \mu_{0i} + \mu_{1i}t$ ($(\mu_{0i},\mu_{1i}) \in \mathbb{R}^2$ are unknown random effects), the model:

$$\tilde{Y}_i(t_{ij}) = X_i^T(t_{ij})\mu_i(t_{ij}) + e_{ci}(t_{ij})$$

can be estimated locally using the kernel weighted local linear random effect regression by

Algorithm 2 Backfitting algorithm for the MVCRE model estimation

1. **Input**: Outcome $(Y_i(t_{ij}))_{i=1,\dots,N_s,j=1,\dots,N_i}$, covariates $(X_i(t_{ij}))_{i=1,\dots,N_s,j=1,\dots,N_i}$, time points $(t_{ij})_{i=1,\dots,N_s,j=1,\dots,N_i}$, number of components C and bandwidths h_f and h_r for estimating the time-varying coefficients and the random effects processes respectively, subject identifiers (noted by $i = 1, \dots, N_s$) and maximum number of iterations max_iter.
2. Initialize the parameters $\theta = (\pi,\ \beta,\ \mu,\ \sigma)$, where $\pi = (\pi_c(t))_{c=1,\dots,C}$, $\beta = (\beta_c(t))_{c=1,\dots,C}$, $\mu = (\mu_i(t))_{i=1,\dots,N_s}$ and $\sigma = (\sigma_c(t))_{c=1,\dots,C}$.
3. Take $\overline{Y}_i(t_{ij}) = Y_i(t_{ij}) - X_i(t_{ij})\mu_i(t_{ij})$. Run k iterations of algorithm 1 to estimate the parameters π, β and σ of the MVC model:

$$\overline{Y}_i(t_{ij}) = \sum_{c=1}^{C} \mathbb{1}_{\{z_i(t_{ij})=c\}}(X_i^T(t_{ij})\beta_c(t_{ij}) + e_{ci}(t_{ij})). \tag{4.3}$$

4. Take $\tilde{Y}_i(t_{ij}) = Y_i(t_{ij}) - (X_i^T(t_{ij})\beta_c(t_{ij}))$ where c is the component allowing for the best fit at time t_{ij}.
5. Get the estimates $\hat{\mu}_i(t)$ of the time-varying random effects in the model:

$$\tilde{Y}_i(t_{ij}) = X_i^T(t_{ij})\mu_i(t_{ij}) + e_{ci}(t_{ij}). \tag{4.4}$$

   Similarly to locally linear regression, we set $\mu_i(t) = \mu_{0i} + \mu_{1i}t$ and then estimate a locally weighted random effects model in equation (4.5) in order to get time-varying estimates $\hat{\mu}_i(t) = \hat{\mu}_{0i} + \hat{\mu}_{1i}t$.
6. Repeat steps 3, 4 and 5 until convergence (*i.e.* we stop the algorithm when the mean squared differences between the estimates of θ from the previous iteration and the current iteration are smaller that 10^{-5}) or until reaching max_iter.
7. **Output**: The estimations of the parameters in θ: $\hat{\beta}_c(t)$, $\hat{\pi}_c(t)$ and $\hat{\sigma}_c(t)$ for $c = 1, \dots, C$ and $\hat{\mu}_i(t)$ for $i = 1, \dots, N_s$.

fitting the weighted mixed effects model:

$$K_{h_r}(t_{ij} - t)^{1/2}\tilde{Y}_{ij} = K_{h_r}(t_{ij} - t)^{1/2}x_{ij}(\mu_{0i} + \mu_{1i}t_{ij}) + e_{cij}. \tag{4.5}$$

In the next section, we will discuss the literature related to the convergence of the backfitting algorithm and propose some directions for choosing the starting values.

4.4 Choice of initial values

In order to estimate the model parameters, we use an iterative method that may eventually converge to local maxima. Several methods used for the choice of initial values in mixture models have been discussed in the literature [142–144]. Although, we did not prove the

convergence of algorithm 2 theoretically, we illustrate our algorithm empirical convergence using simulated data (see section 4.7). In this work, we decided to adopt initialization using grid search [143] because of its general applicability.
We start by generating $(p+1) \times C$ random Gaussian processes with different constant means and different variance functions for the parameters $(\beta_c(t))_{\{c=1,\dots,C\}}$ and C random Gaussian processes for the parameters $(\sigma_c(t))_{\{c=1,\dots,C\}}$. The mixture proportions parameters are initiated with constants equal to $\frac{1}{C}$ for all the components in $1,\dots,C$. Using these initial values, we then estimate the model. The random effects $\mu_i(t)$ are initialized with 0 constant for all t. We repeat this process with several random initialisations. Once the models have been developed, we then keep the model which maximizes the log-likelihood given by:

$$\log(\mathcal{L}) = \sum_{i=1}^{N_s}\sum_{j=1}^{N_i}\log\sum_{c=1}^{C}\pi_c(t_{ij})\ \phi(Y_i(t_{ij})|\mathrm{X}_i^T(t_{ij})\beta_c(t_{ij}) + \mathrm{X}_i^T(t_{ij})\mu_i(t_{ij}), \sigma_c(t_{ij})). \tag{4.6}$$

4.5 Number of mixture components

Determining the number of mixture components is an important step in developing mixture models. Extensive literature exists on this problem. It ranges from using model selection criteria such as Akaike information criterion (AIC) and Bayesian information criterion (BIC) [145] to using Bayesian estimation methods allowing for the adaptation of the number of components to the complexity of the data [146, 147]. In this chapter, we will focus on the former methods using model selection criteria. Huang *et al.* [85] proposed a criterion to estimate the number of mixture components leading to the best model in the case of a mixture of varying-coefficient models with no random effect processes. In their procedure, they proposed a modified BIC that can be written as follows:

$$BIC_{MVC} = -2\log(\mathcal{L}) + (C(p+1)df_\beta + C - 1)\log(N), \tag{4.7}$$

where $df_\beta = \frac{\tau_K}{h_f}|\Omega|(K(0) - \frac{1}{2}\int K^2(t)dt)$, $\tau_K = \frac{K(0)-\frac{1}{2}\int K^2(t)dt}{\int(K(t)-\frac{1}{2}K*K(t))^2dt}$ and Ω is the support of the time covariate t (*i.e.* if $t \in [a,b]$ then $|\Omega| = b-a$). The function $K * K(t)$ is the convolution product of the kernel function K with itself. The quantity df_β represents the complexity of the non-parametric estimation of the fixed effects. The log-likelihood function $\log(\mathcal{L})$ is given in equation (4.6). The theoretical justification of the use of the quantities τ_K and $K(0) - \frac{1}{2}\int K^2(t)dt$ are explained in a paper by Fan *et al.* [148].
Following the same principle as Delattre *et al.* [45] which proposed a BIC for mixed effects models, we generalize the modified BIC, as described in equation (4.7), to the mixture of varying-coefficient models with random effects. We propose to add the term $C(p+1)df_\mu \log(N_s)$ to the previously proposed BIC to account for the dimensionality of random effects. Our added term includes a penalty associated with the number of random effects $q = (p+1)$, the degree of freedom df_μ and the number of subjects N_s. Our modified BIC can be written as:

$$BIC_{MVCRE} = -2\log(\mathcal{L}) + (C(p+1)df_\beta + C - 1)\log(N) + C(p+1)df_\mu\log(N_s), \tag{4.8}$$

where $df_\mu = \frac{\tau_K}{h_r}|\Omega|(K(0) - \frac{1}{2}\int K^2(t)dt)$ is the complexity of the non-parametric estimation of the random effects. As a reminder, N_s is the number of subjects and N is the total number

of observations.
The part of the BIC expression associated with random effects does not depend on the number of mixture components. It can be removed when trying to find the optimal number of components in the mixture. The proposed BIC can also be used in selecting the optimal model (feature selection [149]).
In order to demonstrate the usefulness of our modified BIC, we conducted two simulation studies. The first simulation is used to show that the BIC can help to select from either a model with only time-varying fixed effects [85] or a model with time-varying mixed effects. In other words, our modified BIC can determine whether the inclusion of random effect processes is needed or not. The second simulation is to show that our modified BIC can be used to identify the optimal number of components in the model proposed in this chapter.

4.6 Bandwidth selection and model evaluation

In order to conduct bandwidth selection, which is an important issue in non-parametric models, we choose to use CV. Several methods have been evaluated in the literature [150].
In order to estimate the optimal bandwidth for the time-varying fixed effects model (4.3), Huang *et al.* [85] proposed to choose the bandwidth that maximizes the J-fold CV log-likelihood:

$$CV_1 = \frac{1}{J}\sum_{j=1}^{J}\sum_{l\in D_j}\log\sum_{c=1}^{C}\hat{\pi}_{c(-j)}(t_l)\ \phi(Y_l|X_l^T\hat{\beta}_{c(-j)},\hat{\sigma}_{c(-j)}(t_l)),$$

where J is the number of folds and D_j are the observations in fold j. The quantities $\hat{\pi}_{c(-j)}$, $\hat{\beta}_{c(-j)}$ and $\hat{\pi}_{c(-j)}$ are the estimates of the model parameters obtained without using the data of fold j.
Using data from the same subject for model estimation and model evaluation may lead to underestimation of the true model error for subject-overfitted models. This leads to choosing the bandwidth of an overfitted model over a well adjusted model. Leave-one-subject-out CV was found to be more consistent in repeated measures data [63].
The leave-one-subject-out CV for model (4.3) can be obtained simply by substituting each fold by a one-subject data:

$$CV = \frac{1}{N_s}\sum_{s=1}^{N_s}\sum_{l\in D_s}\log\sum_{c=1}^{C}\hat{\pi}_{c(-s)}(t_l)\ \phi(Y_l|X_l^T\hat{\beta}_{c(-s)},\hat{\sigma}_{c(-s)}(t_l)), \tag{4.9}$$

where N_s is the total number of subjects.
Clearly, separate estimation of the optimal bandwidth for the time-varying fixed effects and the random effects has its limits. It is time consuming because all the combinations between the elements of the set of bandwidths for time-varying fixed effects and the set of bandwidths for random effects should be tested by two separate CV procedures. We have opted to estimated the optimal bandwidth for the time-varying fixed effects using the leave one subject-out cross-validation with a fixed random effects bandwidth and then estimate the model again with the optimal fixed bandwidth using different bandwidths for the random effects and keeping the model yielding the smallest BIC_{MVCRE}.

4.7 Simulations and real data examples

In this section, we illustrated our model using simulated synthetic data and real data. In our simulations, we compared our MVCRE model to the MVC model proposed by Huang *et al.* [85]. We calculated the root mean squared error (RMSE) given by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\beta(t_i) - \hat{\beta}(t_i))^2},$$

between the real known coefficients $\beta(t_i)$ described in the following subsection 4.7.1 and the estimated coefficients $\hat{\beta}(t_i)$ using MVC and MVCRE models. The RMSE is calculated for a set of bandwidths and the optimal bandwidth determined using the proposed CV procedure. Other simulations were conducted to evaluate the proposed modified BIC for model selection purposes. The aim is to show that the modified BIC we propose can be used to choose between the two models. We compare the modified BIC between our model and the MVC model for simulations including random effects and simulations not including random effects. We also conducted simulations in order to evaluate if the BIC is able to identify the number of mixture components (denoted by C) to include in the model.
For the real clinical data example, we used data from the PREDIBACK study, which is an original dataset containing several questionnaire scores of pain intensity, functional disability, depression and quality of life of 200 patients with chronic pain after spinal surgery (CPASS). PREDIBACK study is described in section 4.7.2.

4.7.1 Simulation data

4.7.1.1 Simulation 1: Comparison of the MVC and MVCRE model

In order to compare our model to the MVC model proposed by Huang et al, we used a two-component mixture of varying-coefficient models including random effect processes with two independent variables ($p = 2$). We started by running 100 simulations with sample sizes of $N = 250$ and $N = 500$ with $N_s = 10$ subjects in order to identify the optimal bandwidths from a set of 4 bandwidths ($h_1 = 0.15$, $h_2 = 0.18$, $h_3 = 0.21$ and $h_4 = 0.24$). The bandwidth of random effects was fixed to 0.21 due to the computational burden of the CV procedure. After determining the optimal bandwidths for the time-varying fixed effects using the proposed CV with $J = 5$ folds, we conducted 500 simulations with sample sizes of 250 and 500 using these optimal bandwidths and we illustrated the two models results graphically. The random effects were randomly generated once and fixed for all simulation. We take the following parameters for our simulations:

$$\begin{gathered}
C = 2, \qquad t \sim U(0,1), \qquad X \sim \mathcal{N}((0,0,0)^T,\, I_3) \\
\pi_1(t) = \exp(\tfrac{-t-1}{2}), \qquad \pi_2(t) = 1 - \pi_1(t) \\
\beta_{01}(t) = 1 - t, \qquad \beta_{02}(t) = 4 \\
\beta_{11}(t) = 1 + \cos(2\pi t) + \sin(2\pi t), \qquad \beta_{12}(t) = -1 + \tanh(t) \\
\mu_{ki}(t) = a_{ik0} + a_{ik1}t + a_{ik2}t^2 \text{ for } i \text{ in } \{1,\ldots,10\} \text{ and } k \text{ in } \{0,1\}
\end{gathered}$$

$$\text{where } (a_{00}, a_{01}, a_{02})^T \sim N((0,0,0)^T, diag(0.25, 0.25, 0.25)) \text{ and}$$
$$(a_{10}, a_{11}, a_{12})^T \sim N((0,0,0)^T, diag(0.75, 0.75, 0.75))$$
$$e_{ci} \sim \mathcal{N}(0, 0.5^2) \text{ for } c \text{ in } \{1,2\}$$

$$Y_i(t_{ij}) = \sum_{c=1}^{2} \mathbb{1}_{\{z_i(t_{ij})=c\}}(\beta_{0ci}(t_{ij}) + x_{1i}(t_{ij})\beta_{1ci}(t_{ij}) + e_i^c(t_{ij})),$$

where $\beta_{kci}(t) = \mu_{ki}(t) + \beta_{kc}(t)$ for k in $\{0,1\}$ and c in $\{1,2\}$ (k is the index associated with each independent variable).
The random coefficients a_{kl} (k in $\{0,1\}$ and l in $\{0,1,2\}$) associated with the polynomial random effects were sorted in order to generate more variability over time of the random effects and to observe how the estimates of the MVCRE and MVC model to the data (*i.e.* $a_{kli} < a_{kli+1}$ for i in $\{1,\ldots,N_s\}$). Moreover, we used the Epanechnikov kernel for all our simulations ($K_h(v) = \frac{3}{4}(1-(v/h)^2)$ if $|v/h| < 1$ and 0 otherwise). The number of observations per subject is variable. Subjects were drawn from a discrete uniform distribution. For $N = 250$, the mean number of observations per subject is 25 and the mean number of observations per subject is 50 for $N = 500$. We estimated the time-varying coefficients using both our MVCRE model and the MVC model, which does not include random effect processes. This comparison is conducted in order to illustrate the added value of including time-varying random effects in the model. We have chosen to include several types of coefficient functions ($\beta_{11}(t)$ has high curvature, $\beta_{01}(t)$ is linear, $\beta_{02}(t)$ is constant, *etc*). An illustration of the proposed simulation can be found in Figure . This allows us to illustrate the limits of our proposed model regarding the use of kernel non-parametric estimation. Table 4.1 shows the estimated RMSEs obtained from the estimates using our MVCRE model and from the estimates using the MVC model without random effects [85]. We obtain smaller RMSEs for both $N = 250$ and $N = 500$ simulations using our model with time-varying random effects. The MVCRE model RMSEs are smaller than one standard deviation below the mean of RMSEs obtained using the MVC model. These results suggest that our model yields better estimates of the true effects compared to the MVC model.
After determining the best bandwidths for our simulations, we conducted 500 new simulations using these optimal bandwidths ($h = 0.18$ for the MVCRE model and $h = 0.24$ for the MVC model). We plotted the estimated time-varying coefficients for the two mixture components and their 95% confidence intervals for N=250 (Figure 4.2) and $N = 500$ (Figure 4.3). We can see that our model gives mean estimates which are closer to the true coefficients. This is due to the fact that our model estimates the coefficients by pooling the subject-specific coefficients (random effects) instead of estimating overall coefficients by supposing that all data points are independent. This can be observed in Figure 4.4. This figure shows the true fit of $Y(t)$ using $X(t)$ and the true coefficients and the subject-specific fit of $Y_i(t)$ using $X_i(t)$ and the subject effect $\beta_i(t)$. As can be observed in Figure 4.4, the MVC model proposed by Huang et al tries to find the best fit based on the full data without considering intra-subject differences. On the other hand, our MVCRE model finds the best fit by averaging the estimated random effects, which leads to a better, less biased estimation of the true effects.
When comparing the results of the simulations obtained using $N = 250$ and $N = 500$ in Figure 4.2 and 4.3, we can observe that both bias and the width of confidence intervals

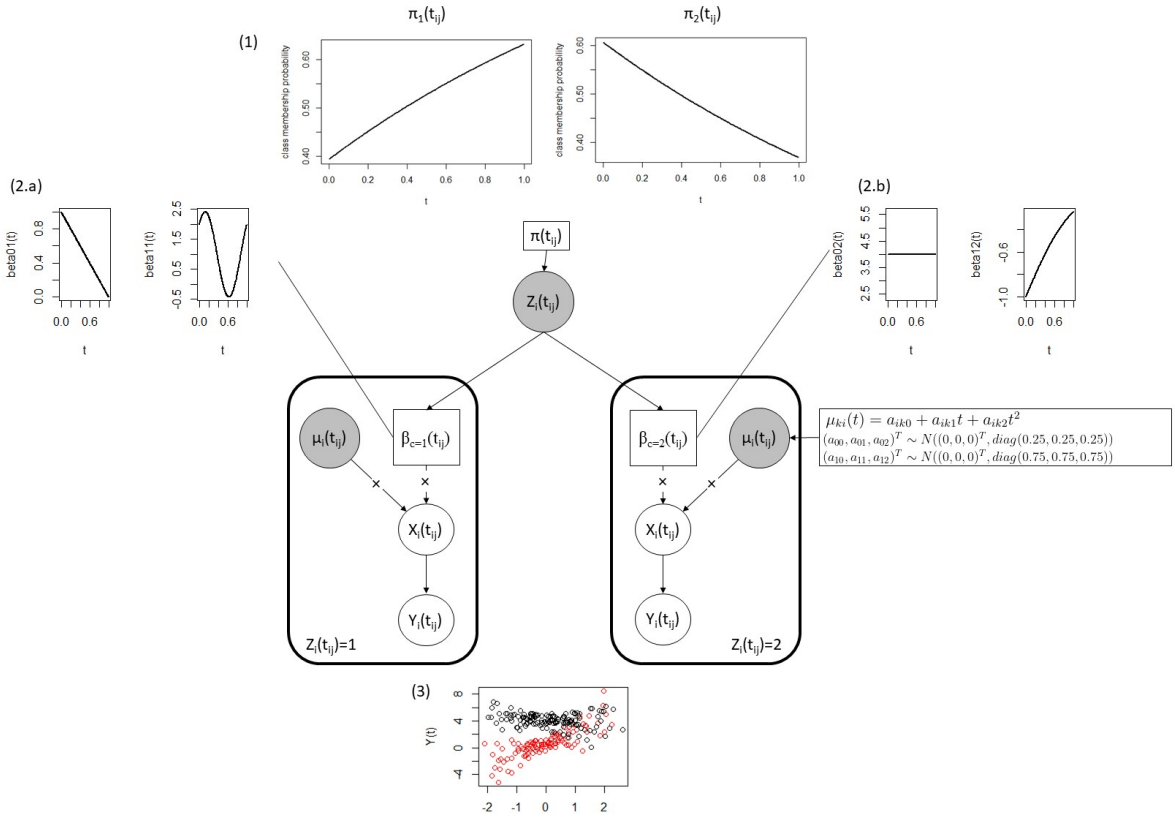

FIGURE 4.1 – A graphical model representation of the simulated data and the different parameters to estimate. The quantities in a white square represent the parameters which should be estimated. The quantities in a white circle represent the observed variables and the quantities in a grey circle represent unknown latent variables. The plot (1) represents the simulated class membership probabilities for the two classes. The plots (2.a) and (2.b) are the simulated trajectories of the time-varying regression coefficients for class 1 and class 2 respectively. The scatter plot (3) illustrates the simulated outcome $Y(t)$ as a function of the dependent variable $X(t)$ for each class (Class 1 in red and class 2 in black), without including the noise $e(t)$.

become smaller with a larger sample size suggesting more precise estimates of the true coefficients. We can also see that the MVCRE estimates of coefficients with low curvature (constant $\beta_{02}(t)$ and linear $\beta_{01}(t)$ coefficients) are more accurate than the estimates of coefficients with high curvature (*e.g.* $\beta_{11}(t)$). This is due to the use of a common bandwidth for all coefficients. Coefficients with higher curvature generally need higher bandwidths in order to capture the changes in the coefficient. On the other hand, lower curvature coefficients need lower bandwidths in order to avoid unnecessary variability. This can be resolved by estimating an optimal bandwidth for each coefficient [71, 151]. The downside is that this method leads to a computationally intensive CV optimal bandwidth estimation.

TABLE 4.1 – Quantitative comparison of our model and the mixture of varying-coefficient models. The two models were compared using the metric RMSE and its standard deviation for $h_1 = 0.15$, $h_2 = 0.18$, $h_3 = 0.21$ and $h_4 = 0.24$. The values highlighted in bold font are those associated with the bandwidth yielding the smallest RMSE for a given coefficient. The optimal bandwidth for the time-varying fixed effects was determined only for $N = 500$. RMSE (standard deviation) were estimated using bootstrap on 100 simulations. The optimal bandwidth was determined using the proposed CV with $J = 5$.

Coefficient	**Sample size**	**MVCRE**			
		h_1	$h_2 = h_{opt}$	h_3	h_4
β_{01}	N=250	0.36 (0.34)	0.31 (0.28)	0.28 (0.10)	**0.27 (0.10)**
β_{11}		**0.32 (0.18)**	0.55 (0.24)	0.41 (0.19)	0.52 (0.22)
β_{02}		0.73 (0.46)	0.89 (0.37)	**0.71 (0.24)**	0.79 (0.26)
β_{12}		0.41 (0.13)	0.42 (0.13)	**0.34 (0.10)**	0.35 (0.10)
β_{01}	N=500	0.10 (0.03)	0.08 (0.02)	0.07 (0.02)	**0.07 (0.02)**
β_{11}		**0.26 (0.03)**	0.29 (0.03)	0.33 (0.03)	0.38 (0.03)
β_{02}		0.22 (0.08)	0.13 (0.06)	0.11 (0.04)	**0.09 (0.03)**
β_{12}		0.09 (0.03)	0.09 (0.03)	0.09 (0.03)	**0.09 (0.03)**
		MVC			
		h_1	h_2	h_3	$h_4 = h_{opt}$
β_{01}	N=250	1.49 (0.10)	**1.39 (0.16)**	1.48 (0.07)	1.46 (0.08)
β_{11}		1.11 (0.11)	1.16 (0.12)	**1.11 (0.10)**	1.15 (0.11)
β_{02}		2.12 (0.09)	2.20 (0.15)	2.09 (0.08)	**2.06 (0.09)**
β_{12}		1.48 (0.08)	1.52 (0.09)	1.44 (0.07)	**1.40 (0.07)**
β_{01}	N=500	1.42 (0.07)	1.40 (0.06)	1.40 (0.06)	**1.38 (0.06)**
β_{11}		1.13 (0.14)	0.98 (0.15)	**0.96 (0.16)**	0.99 (0.17)
β_{02}		2.07 (0.09)	2.10 (0.07)	2.04 (0.07)	**2.01 (0.06)**
β_{12}		**1.50 (0.22)**	1.74 (0.24)	1.75 (0.26)	1.67 (0.26)

In addition, we can observe a common limit of kernel estimators, which is the boundary problem. In order to correct estimation near edges, we plan to extend the present work by considering Bernstein polynomials rather than kernels [152]. It is also possible to consider Lagrange polynomials with Tchebychev-Gauss points [153]. Other methods of boundary correction such as the jackknife and translation bootstrap can be found in the literature [154]. The empirical convergence of our estimates using the backfitting EM algorithm for sample size $N = 500$ can be found in Figure 4.5. We can see that after a burn-in period of approximately $M = 15$ iterations, the RMSE values become quite stable after reaching their minimum value. The conditional log-likelihood given in equation 4.6 increases at each iteration until it reaches a local maximum. This is due to the fact that the EM algorithm monotonically increase the conditional log-likelihood (4.2) at each iteration until reaching a local maximum.

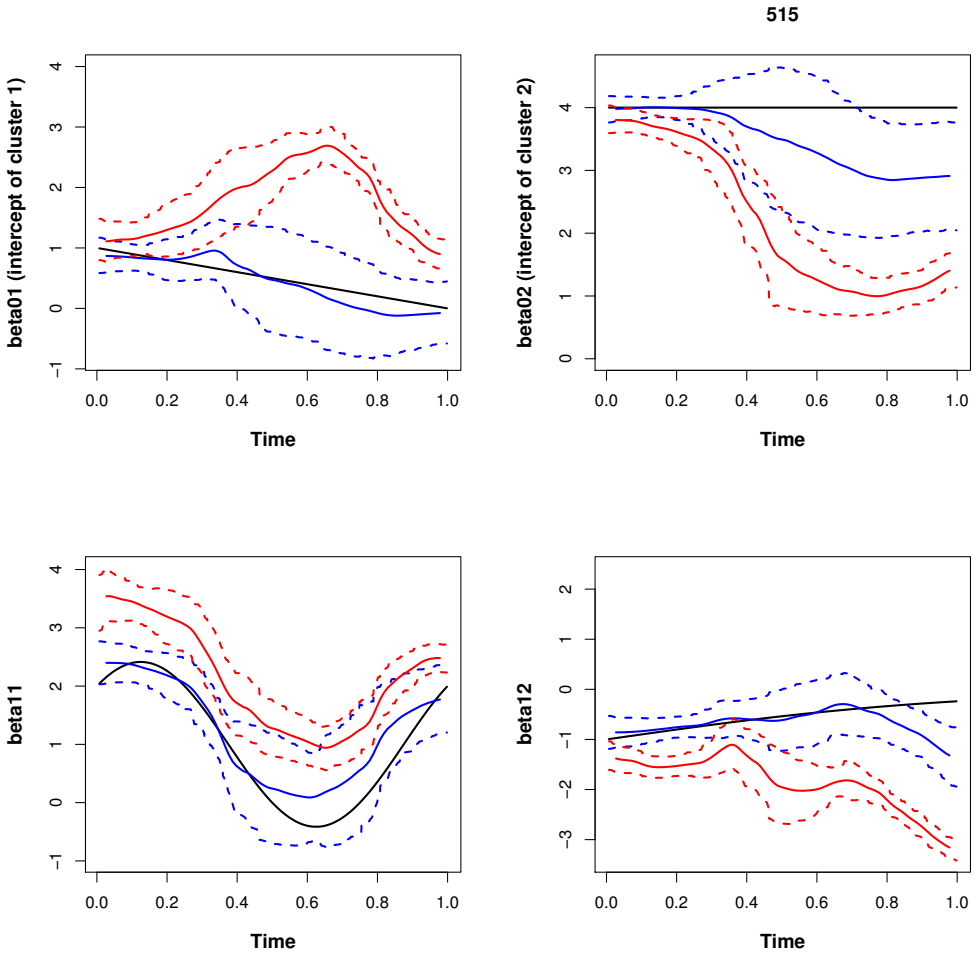

FIGURE 4.2 – The model coefficients and their 95% confidence interval estimated using 500 simulations with a sample size of N=250. Black curves represent the real coefficients. The solid curves in red represent the mean estimated time-varying coefficients obtained using the MVC model with no random effects while the dashed red curves are their 95% confidence intervals. The solid curves in blue are the mean estimated time-varying coefficients obtained using the model considering random effects. The dashed blue curves are their 95% confidence intervals. The first column contains the coefficients of the first mixture component ($\beta_{01}(t)$, $\beta_{11}(t)$ and $\beta_{21}(t)$) while the second column contains the coefficients obtained for the second component ($\beta_{02}(t)$, $\beta_{12}(t)$ and $\beta_{22}(t)$).

4.7.1.2 Simulation 2: Use of the modified BIC as a model selection criterion

In this second simulations, our aim was to compare the MVC model and the MVCRE model when the variance of the random effects is small. In the classical mixed effects framework, when the variability of the random effects is small and the data are balanced, the mixed effects model and the multiple regression model yield similar fixed effect estimates. We also aimed to show that even when the subject-level variation is small, the proposed BIC_{MVCRE} still manages to identify the true model regarding the inclusion of

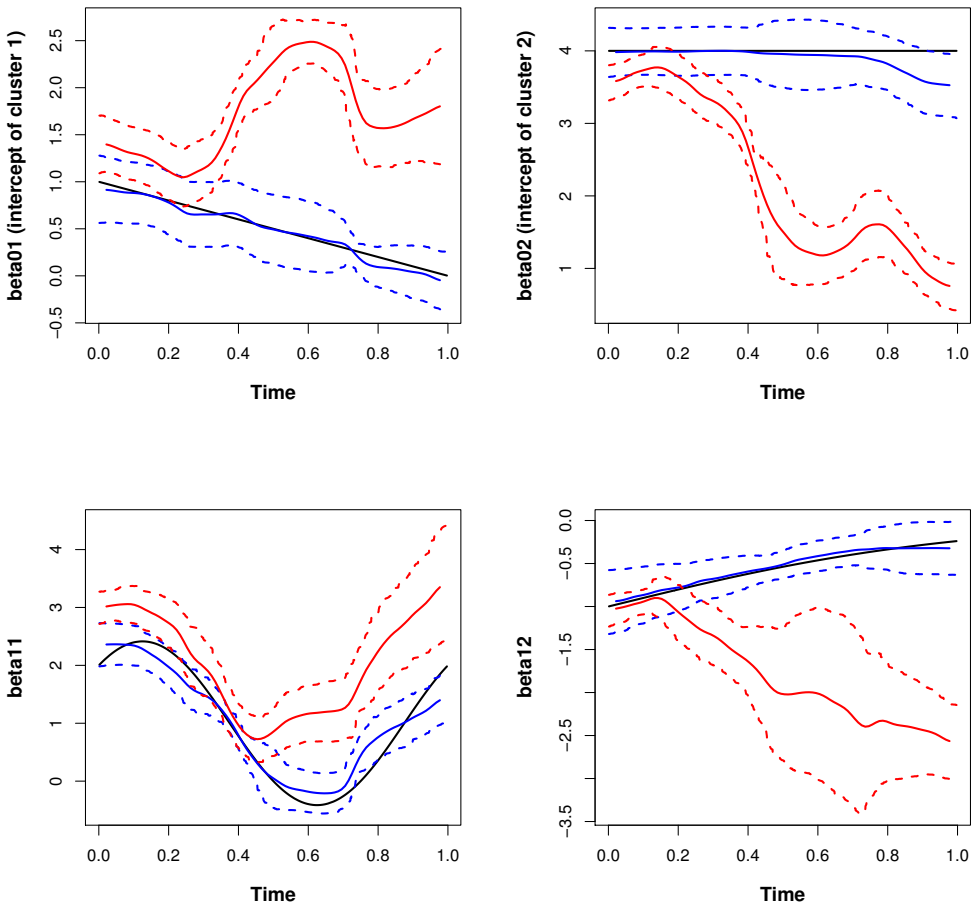

FIGURE 4.3 – The model coefficients and their 95% confidence interval estimated using 500 simulations with a sample size of N=500. Black curves represent the real coefficients. The solid curves in red represent the mean estimated time-varying coefficients obtained using the MVC model with no random effects while the dashed red curves are their 95% confidence intervals. The solid curves in blue are the mean estimated time-varying coefficients obtained using the model considering random effects. The dashed blue curves are their 95% confidence intervals. The first column contains the coefficients of the first mixture component ($\beta_{01}(t)$, $\beta_{11}(t)$ and $\beta_{21}(t)$ while the second column contains the coefficients obtained for the second component.

random effects and the number of components.
In this simulation, the following parameters were used:

$$C = 2, \qquad t \sim U(0,1), \qquad X \sim \mathcal{N}((0,0,0)^T,\ I_3)$$
$$\pi_1(t) = \exp(\tfrac{-t-1}{2}), \qquad \pi_2(t) = 1 - \pi_1(t)$$
$$\beta_{01}(t) = 1 - t, \qquad \beta_{02}(t) = 4$$
$$\beta_{11}(t) = 1 + \cos(2\pi t) + \sin(2\pi t), \qquad \beta_{12}(t) = 1 - 2t^2$$
$$\beta_{21}(t) = \log(t^2), \qquad \beta_{22}(t) = \tanh(t)$$
$$\mu_{ki}(t) = a_{ik0} + a_{ik1}t + a_{ik2}t^2 \text{ for } i \text{ in } \{1,\dots,10\} \text{ and } k \text{ in } \{0,1,2,3\}$$
$$\text{where } (a_{k0}, a_{k1}, a_{k2})^T \sim N((0,0,0)^T, diag(0.5, 0.25, 0.1))$$

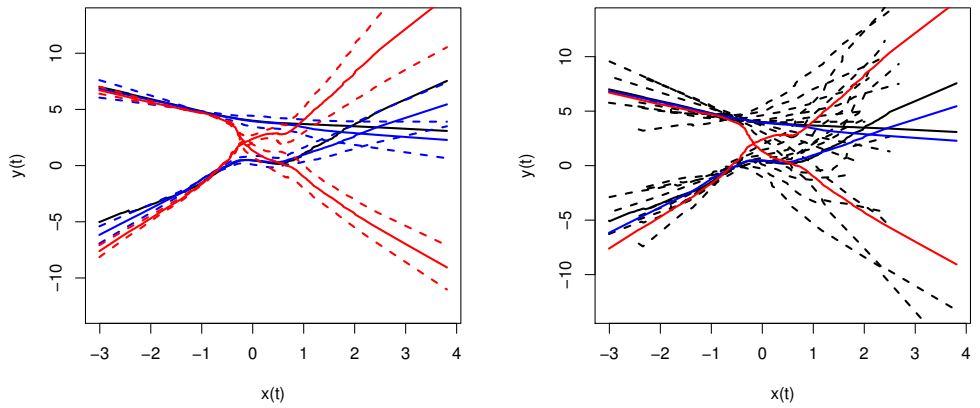

FIGURE 4.4 – The left figure shows the fit of $Y(t)$ obtained from $X(t)$. The black solid lines are the fit obtained using the true coefficients of the two classes. The blue lines are the fit obtained using the MVCRE estimated coefficients for the two classes while the red lines is the fit obtained using the MVC estimated coefficients for the two classes. The dashed blue lines and dashed red lines are the confidence intervals obtained using the MVCRE model and MVC model respectively. We can observe that the MVCRE model fits better to the true data compared to the MVC model. The right figure also includes the fits obtained using the true random effects for each subject (dashed black lines). We can see that the MVCRE model and the MVC model estimates are different regarding their aim since the MVCRE model finds the global effects by "pooling" the random effects while the MVC model just maximizes the global likelihood under the assumption that all points are independent.

$$e_{ci} \sim \mathcal{N}(0, 0.25^2) \text{ for } c \text{ in } \{1, 2\}$$

$$Y_i(t_{ij}) = \sum_{c=1}^{2} \mathbb{1}_{\{z_i(t_{ij})=c\}}(\beta_{0ci}(t_{ij}) + x_{1i}(t_{ij})\beta_{1ci}(t_{ij}) + x_{2i}(t_{ij})\beta_{2ci}(t_{ij}) + e_i^c(t_{ij})),$$

Similar to the first simulation, we conducted 100 simulations with $N = 500$ and $N_s = 10$ subjects to identify the optimal bandwidth for $h \in \{0.12, 0.15, 0.21, 0.24\}$ using the proposed CV procedure for the MVCRE and MVC models. The optimal bandwidth for the random effects was fixed to 0.25. The optimal bandwidths were then used to conduct 500 simulations for the two models. The estimated coefficients for both the MVCRE and the MVC models are illustrated in Figure 4.6.

We can observe in Figure 4.6 that, in this simulation, the biases of the coefficients obtained by the MVCRE model and the MVC model are similar but that the 95% confidence intervals width differs between the two models. In the first simulation where the variability of the random effects was larger and data were unbalanced, we found considerable differences in bias between the two models. When data are balanced, models with and without random effects yield similar coefficient estimates but with different standard errors. Indeed, we can observe that we obtain narrower 95% confidence intervals with the MVCRE model with the same number of simulations. This comes from the standard errors of the coefficients estimated using the MVCRE model being smaller than those obtained using the MVC model. In this simulation, although the two models had similar biases, the MVCRE model gives estimates that show less variability around the estimated coefficients.

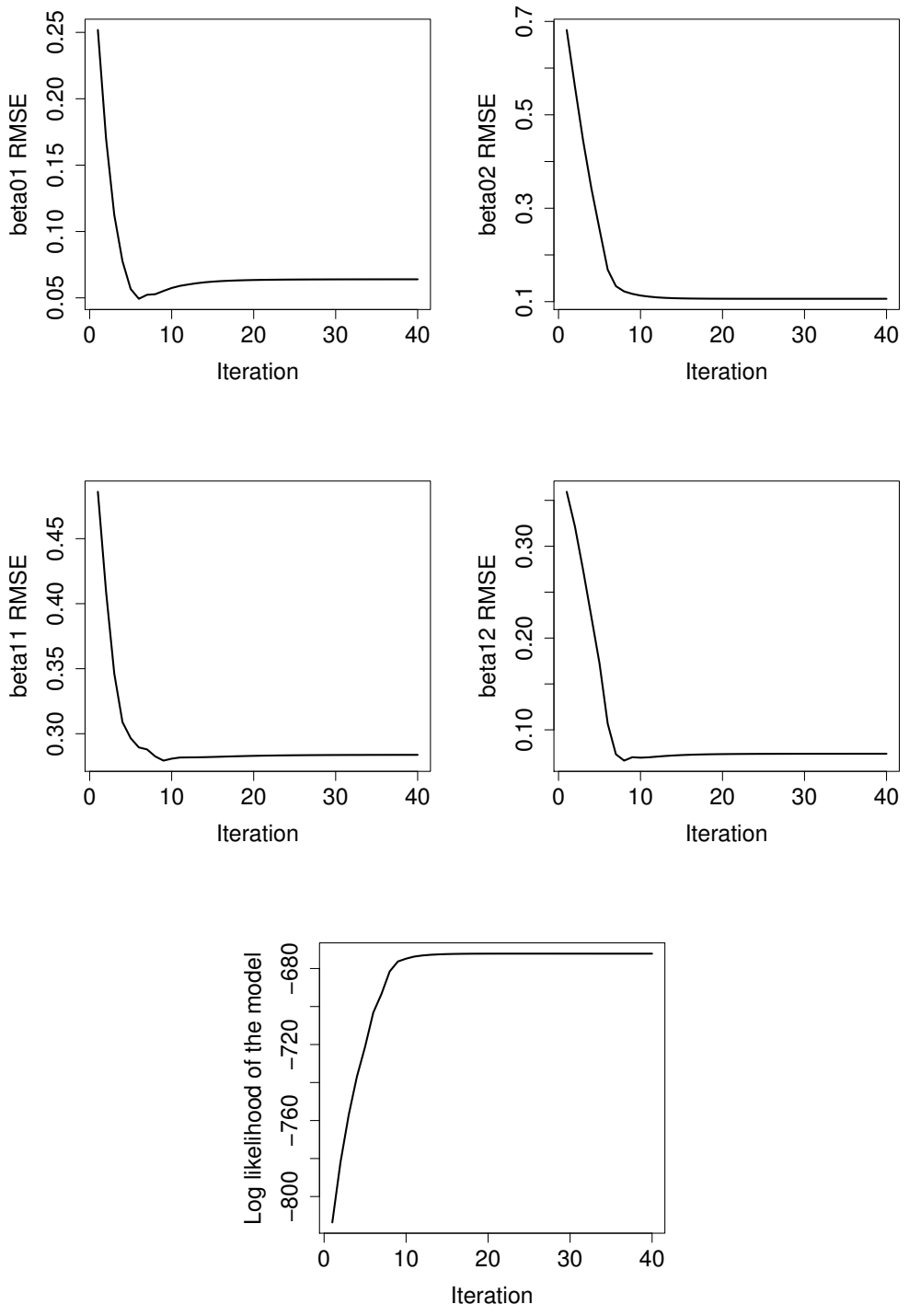

FIGURE 4.5 – The four upper figures represent the RMSE associated with each coefficient at each iteration of the backfitting EM algorithm. The fifth figure represents the conditional log-likelihood function in 4.6.

TABLE 4.2 – Contingency table between the selected models based on the modified BIC, and the true simulated model (N=500). The diagonal elements are the correctly identified models.

		Minimal BIC		
		MVCRE	MVC	Total
Correctly specified true model	MVCRE	500 (100%)	0 (0%)	500
	MVC	43 (8.6%)	457 (91.4%)	500

In the following simulation, our aim was to evaluate if the modified BIC is able to identify if the random effects should be included in the model. In practical terms, we wanted to assess the ability of the modified BIC to determine what model should be used: the MVCRE model we proposed in this chapter or the MVC model. We conducted 500 simulations with $N = 500$ observations each, following the same configurations as the previous simulations in subsection 4.7.1.2 except that the random effects $\mu_i(t)$ were set to 0 for all t and for all subjects. In other terms, we did 500 simulations where no random effects were considered. We calculated the BIC_{MVC} and BIC_{MVCRE}, given in equations (4.7) and (4.8), for the 500 simulations with no random effects and the 500 simulations with random effects. The optimal bandwidths identified previously in section were used for these simulation. We considered the model MVCRE to be better than MVC if the BIC_{MVCRE} was smaller than BIC_{MVC}. We reported the contingency table (Table 4.2) between the selected model (the model with the minimal BIC) and the correctly specified true model. Our simulations show that the proposed modified BIC can help to correctly identify if the random effects should be included or not in the model. When the correctly specified model was the MVCRE model, the BIC selected the MVCRE model for 100% of our 500 simulation. On the other hand, when the true model was the MVC model, the BIC selected the MVC model for 91.4% of our simulations.

4.7.1.3 Simulation 3: Use of the modified BIC for identifying the optimal number of components

In this simulation, we show that the proposed BIC_{MVCRE} can also be used to determine the number of components C to include in the proposed mixture model. For this, we used the same configurations as the previous simulations in subsection 4.7.1.2.
We conducted 100 simulations with $N = 250$ observations. For each simulation, we model the simulated data using the MVCRE model with different numbers of mixture components. We started by simulating data with $C = 1$ (*i.e.* taking only the first component with the parameters β_{01} and β_{11} and β_{21} and σ_1) and $C = 2$ (*i.e.* adding the second component with the parameters β_{02} and β_{12} and β_{22} and σ_2 for $C = 2$). We then estimated these parameters using the proposed MVCRE model with different number of mixture components ($C = 1, 2, 3$ and 4). We compared the BIC_{MVCRE} for these four models and we reported the contingency table between the number of components resulting in the minimal BIC_{MVCRE} and the true number of components. The results are shown in Table 4.3.

We can see in Table 4.3 that when the true number of mixture components is equal to 1, the BIC_{MVCRE} was able to identify the true model for 100% of our simulations. When the true number of components is equal to 2, our selection criterion was also able to identify the true

TABLE 4.3 – Frequencies of selected number of mixture components based on the modified BIC, for each true number of components.

		Minimal BIC_{MVCRE}				
		$C = 1$	$C = 2$	$C = 3$	$C = 4$	Total
Correctly specified true model	C=1	100	0	0	0	100
	C=2	8	86	6	0	100

number of components (the true model was identified for 86% of our simulations). These results suggest that our modified BIC_{MVCRE} can be used to identify with good precision the true number of mixture components that should be included in the MVCRE model.

4.7.2 Real data example

4.7.2.1 Real data context

In our real data example, we consider the PREDIBACK study, which comprises an original dataset consisting of clinical, sociodemographic and cognitivo-behavioral evaluation of patients with Chronic Pain After Spinal Surgery (CPASS). CPASS is characterized by persistent pain following one or several spine surgeries [98]. This disabling condition is observed in approximately 20% of patients who having undergone a spinal surgery [14].
CPASS impacts the patient functional capacity, psychological state, daily activities and overall quality of life [30, 123]. Patient's evaluation in practice remains mainly focused on pain intensity [23]. Patients' care needs to include the management of pain dimensions other than pain perception such as functional incapacity and psychological distress [155]. The impact of functional incapacity, pain perception and psychological distress on quality of life can vary between patients and over time but this variability is currently not considered in quality of life evaluation. In this chapter, we used our MVCRE model to identify the temporal change in the impact of pain intensity, functional disability and psychological distress on the quality of life of CPASS patients.
Our aim was to identify clusters of patients based on the impact of pain intensity, functional disability and depression on their health-related quality of life, in order to:

— understand the longitudinal links between pain perception, disability and psychological distress and patient quality of life ;
— potentially propose personalized evaluation according to the patient's temporal stage in the pain trajectory.

4.7.2.2 Data and model description

For this real data example, we used the PREDIBACK study data. It is a prospective longitudinal study, including 200 CPASS patients. Patients were included in 5 French pain management centers (Niort, La Rochelle, Angoulême, Bressuire and Poitiers). The study was registered at clinicaltrials.gov/ct2/show/NCT02964130 and was approved by the Ethics Committee (CPP Ouest III).
We have data from 200 patients with a maximum of 5 observations per patient over 1 year (at inclusion in the study, at 3 months, 6 months, 9 months and 12 months). There are 4

variables per patient, 3 independent variables (scores) that assess pain intensity, depression, and functional disability. The dependent variable is the quality of life score. We wanted to study the evolution over time of the impact of the independent variables on quality of life. Among the 200 patients, 2 did not complete the questionnaires and were removed from the analysis, 26 (13.1%) patients had only 1 visit, 17 (8.6%) had 2 visits, 15 (7.6%) patients had 3 visits, 26 (13.1%) of the patients had 4 visits, and 114 (57.6%) of the patients had 5 visits, for a total of $N = 779$ observations. In the analysis, the date of onset of pain was considered as time t0. Just to clarify, pain onset might occur several years before the patient is included in the study. Data from pain onset until the inclusion in the study are considered missing. The study population includes patients who have had pain for years and for whom no treatment provided adequate pain relief or who were treated but whose pain returned after a period of relief. The mean duration between pain onset (t_0) and the first study visit was 18 years (SD = 13.9 years) (minimum = 6 months and maximum = 46 years).

We have patients for whom we have data 6 months after the onset of pain and other patients for whom we only have data several years after the onset of pain. The study data are unbalanced because the patients are assessed at different times in their pathway and the number of observations per patient is different from patient to patient (from 1 observation to 5 observations). This makes our model more relevant because the inclusion of random effects makes the model more robust to unbalanced data.

Health-related quality of life was evaluated using the EuroQol 5 Dimensions 5-Level (EQ-5D) questionnaire [108], which includes 5 items assessing pain intensity, mobility, self-care, daily activities and anxiety/depression levels. The items consist of a 5-level Likert scale indicating the severity of pain impact on each of the 5 dimensions. The item answers are converted to a score ranging from -0.53 (worse than death quality of life) to 1 (best quality of life possible) based on a dataset of representative French population. Pain intensity was evaluated using a Numeric Pain Rating Scale (NPRS) [112], which is a scale ranging from 0 (no pain) to 10 (worst pain possible). Function disability was assessed using the Oswestry Disability Index (ODI) questionnaire [113]. This questionnaire includes 10 items measuring the impact of pain on personal care, lifting, walking, sitting, sleeping, sexual and social life and travelling. Each item is a Likert scale ranging from 0 (completely able to perform the task) to 5 (unable to perform the task). The final score is calculated as the sum of item scores divided by the number of completed items. Higher scores (greater than 60%) indicate major disability. Finally, depression was evaluated using the Hospital Anxiety and Depression Scale (HADS) [116]. This questionnaire contains 14 4-level Likert scales. 7 items are associated with depression symptoms and the remaining 7 items are associated with anxiety symptoms. The HADS depression and anxiety scores range from 0 to 24 where higher scores indicate higher depression and anxiety levels. We developed an MVCRE model with 3 independent variables, which can be written as follows:

$$\begin{aligned} Y_i(t_{ij}) = \sum_{c=1}^{C} \mathbb{1}_{\{z_i(t_{ij})=c\}} (\beta_{0ci}(t_{ij}) + X_{1i}(t_{ij})\beta_{1ci}(t_{ij}) + X_{2i}(t_{ij})\beta_{2ci}(t_{ij}) \\ + X_{3i}(t_{ij})\beta_{3ci}(t_{ij}) + e_{ci}(t_{ij})), \end{aligned} \tag{4.10}$$

where $Y_i(t_{ij})$ is the EQ-5D score i at time t_{ij} in $1, \ldots, 360$ and $\beta_{0ci}(t_{ij}) = \beta_{0c}(t_{ij}) + \mu_{0i}(t_{ij})$, where $\beta_{0c}(t_{ij})$ is the time-varying intercept and $\mu_{0i}(t_{ij})$ are patient specific random intercepts. The quantities $X_{1i}(t_{ij})$, $X_{2i}(t_{ij})$ and $X_{3i}(t_{ij})$ represent the NPRS score, the ODI score and the HADS depression score respectively. The coefficients $\beta_{kci}(t_{ij}) = \beta_{kc}(t_{ij}) + \mu_{ki}(t_{ij})$ are the fixed and random effects associated with the variables $X_{ki}(t_{ij})$ for $k \in 1, 2, 3$. The time variable t_{ij} represents the time in months between initial pain onset and the evaluation visit. All the variables were normalized to a 0 mean and unit variance. The time variable t was scaled to a $[0, 1]$ interval.
Since higher NPRS, ODI and HADS depression scores represent greater pain, disability and depression level and higher EQ-5D score represent greater quality of life, the model coefficients should be interpreted as more significant when they are more below 0 or when they are higher in absolute value.
The number of mixture components C was identified using the BIC_{MVCRE} criterion. For each treatment modality, the proposed model was developed using $C = 1$, 2 and 3 components. The model having the minimal BIC_{MVCRE} was selected.
The optimal bandwidth for the time-varying fixed effects h_f were selected using the proposed CV procedure. The bandwidth for random effects was set at $h_r = 0.4$. The bandwidths $h_f \in \{0.25, 0.30, 0.35\}$ were tested.

4.7.2.3 Results

The number of mixture components with the minimal BIC_{MVCRE} was $C = 2$ ($BIC_{MVCRE} = 263.9$ for $C = 1$, $BIC_{MVCRE} = 232.6$ for $C = 2$ and $BIC_{MVCRE} = 411.0$ for $C = 3$). The difference between the BIC_{MVCRE} for $C = 1$ and $C = 2$ exceeds 10 suggesting that the model with two components is considerably better. The optimal bandwidth for the time-varying fixed effects was $h_f = 0.35$ which resulted in a CV log-likelihood of 5.2. The estimated time-varying effects for each independent variable can be found in Figure 4.7.
We can see in Figure 4.7 that, for both classes of patients, the coefficients associated with NPRS obtained using the MVCRE model increased towards 0, but that the class 2 patients increased more rapidly and reached very low values (in absolute value). For class 2 patients, the NPRS coefficient went from $\beta_{12}(0) = -0.116$ to $\beta_{12}(1) = -0.017$. This suggests that the effect of pain intensity (NPRS) on quality of life (EQ-5D index) is greater when chronic pain is relatively recent (6 months) than when pain has been present for 46 years ($t = 1$ is equivalent to 46 years), specially for class 2 patients. On the other hand, the effect of ODI on quality of life becomes greater for class 2 patients while its effect is quasi-constant for class 1 patient. The coefficient associated with ODI for class 1 patients was $\beta_{21}(0) = -0.492$ at $t = 0$ and $\beta_{21}(1) = -0.498$ at $t = 1$ showing constant evolution over time. For class 2 patients, the coefficient associated with ODI was equal to $\beta_{22}(0) = -0.436$ at $t = 0$ then decreased to $\beta_{22}(0.64) = -0.630$ at $t = 0.64$, which is equivalent to a 29.5 years duration of pain. The coefficient then increases to $\beta_{22}(1) = -0.488$. This indicates that while functional activity is important for both classes, for class 2 patients, its impact becomes more important for patients with pain for approximately 29 years. Following 29 years of pain, the impact of functional disability decreases until it reaches a similar impact to patients with recent pain after spinal surgery. Finally, regarding the HADS depression score, its coefficient on EQ-5D score decreased from $\beta_{31}(0) = -0.204$ to $\beta_{31}(1) = -0.279$ for class 1 patients while

for class 2 patients, HADS depression score coefficient increased from $\beta_{32}(0) = -0.215$ to $\beta_{32}(0.5) = -0.129$ ($t = 0.5$ is equivalent to 23 years of pain). After 23 years, the HADS score coefficient decreased to $\beta_{32}(1) = -0.292$.
The clinical interpretations of this study are discussed in section 4.8.1.

4.8 Discussion

In this chapter, we developed a mixture of random varying-coefficient models that can be used for time-dependent clustering of the relationship between a continuous time-varying response variable and several time-varying explanatory variables while taking into account intra-subject correlations. Our simulations show that by including intra subject-effects we are able to obtain better estimates than when using the MVC model. We also proposed a modified BIC criterion that can be used to identify the number of classes and also to select the correct model between the mixture of varying-coefficient model without random effects and the mixture of varying-coefficient model with random effects processes proposed in this chapter.

4.8.1 PREDIBACK findings discussion

A real data example we used in this chapter included data from the PREDIBACK study. In this example, we showed that the impact of pain intensity, functional disabiliy and psychological distress on the quality of life of patients with back and leg pain after spinal surgery varied over time. When pain was recent, the quality of life of two classes was impacted by disability and depression similarly but the impact of disability was more important. On the other hand, the impact of pain intensity was greater for class 1 patients compared to class 2 patients. In addition, the impact of pain intensity on quality of life also becomes less significant over time for both classes but this decrease in the coefficient absolute value was more observed in class 2 patients. This decrease in the impact of pain intensity on patients with chronic pain has been observed in a study using brain imaging data [20]. The authors demonstrated that there were significant differences brain activity between patients with acute/subacute pain (pain for less than 3-months) and patients with chronic pain. They found that brain activity in patients with acute pain was focused in the area associated with acute pain perception while for chronic pain patients, this activity shifted to the area associated with emotional/affect perception and the activity associated with pain perception decreased significantly for chronic pain patients. In addition, in a review by Ballantyne & Sullivan published in the New England Journal of Medicine on the use of pain intensity as a metric for evaluating chronic pain [19], they suggested that pain intensity does not reflect patients health state in chronic pain patients since it does not reflect the suffering associated with the suffering and burden due to continuous pain. This might also explain why the increase (in absolute value) in the impact of functional disability over time for class 2 patients. The disability generated by chronic pain leads to a high financial and social burden which are associated with a reduced quality of life [156]. The impact of functional disability on quality of life in class 2 patients decreased after 29.5 years duration of pain. This is consistent with the findings by Wettstein *et al.* [157] in a study evaluating the impact of age on the relationship between disability and well-being in

chronic back pain patients. They found that older patients had better mental health and well-being although they showed higher disability than younger patients, suggesting that their well-being is less impacted by disability. In our results, patients with a 29.5 years pain duration are more likely to be patients of higher ages. These behaviors were not observed in class 1 patients. For these patients, Although a small decrease in the impact of pain intensity on quality of life was observed, the impact of functional disability remained constant over time. However, they showed a increase (in absolute value) in the impact of depression on quality of life. For these patients quality of life is constantly impacted by functional disability but, over time, the psychological/emotional component of pain overtakes the sensory component (pain intensity) similarly to the phenomenon described in the work of Hashmi *et al.* [20].

4.8.2 Strengths, limitations and perspectives

The strength of this model resides in the inclusion of time-varying random effects, which are usually present in clinical data. To our knowledge, the model we propose is the most general currently available model for estimating time-varying, heterogeneous, intra and inter-subject effects in longitudinal data analysis. However, this model is not free of limitations.
Further evaluations of its asymptotic properties and convergence conditions are needed in order to assess the assumptions behind its applicability.
Our model can be extended in future works. In this chapter, the procedures used to test whether a coefficient varies with time or is constant over time were not addressed. Semi-parametric models can be used when some of the coefficients are constant over time. This would probably reduce the risk of over-fitting and estimation time and would increase the interpretability of the model. A generalized likelihood ratio test for non-parametric models with no random effects has been proposed in the literature [85, 148]. It has been shown to verify Wilks phenomena. In the case of linear parametric mixed-effect models, several works studied the asymptotic properties of the likelihood ratio test. It was found that the asymptotic property of Wilk does not hold when random effects are included in the model [158]. Several authors worked on generalizations of the likelihood ratio test for mixed effects models [159, 160]. Future studies are needed to determine the asymptotic proprieties of the likelihood ratio test for random varying-coefficient models.
Another extension that could be explored would consider multiple outcomes as response variables, where several indicators could be combined into a single variable/latent process, similar to the factor analysis and structural equation modeling frameworks.
We are currently working on an `R` package containing the models studied in this chapter.

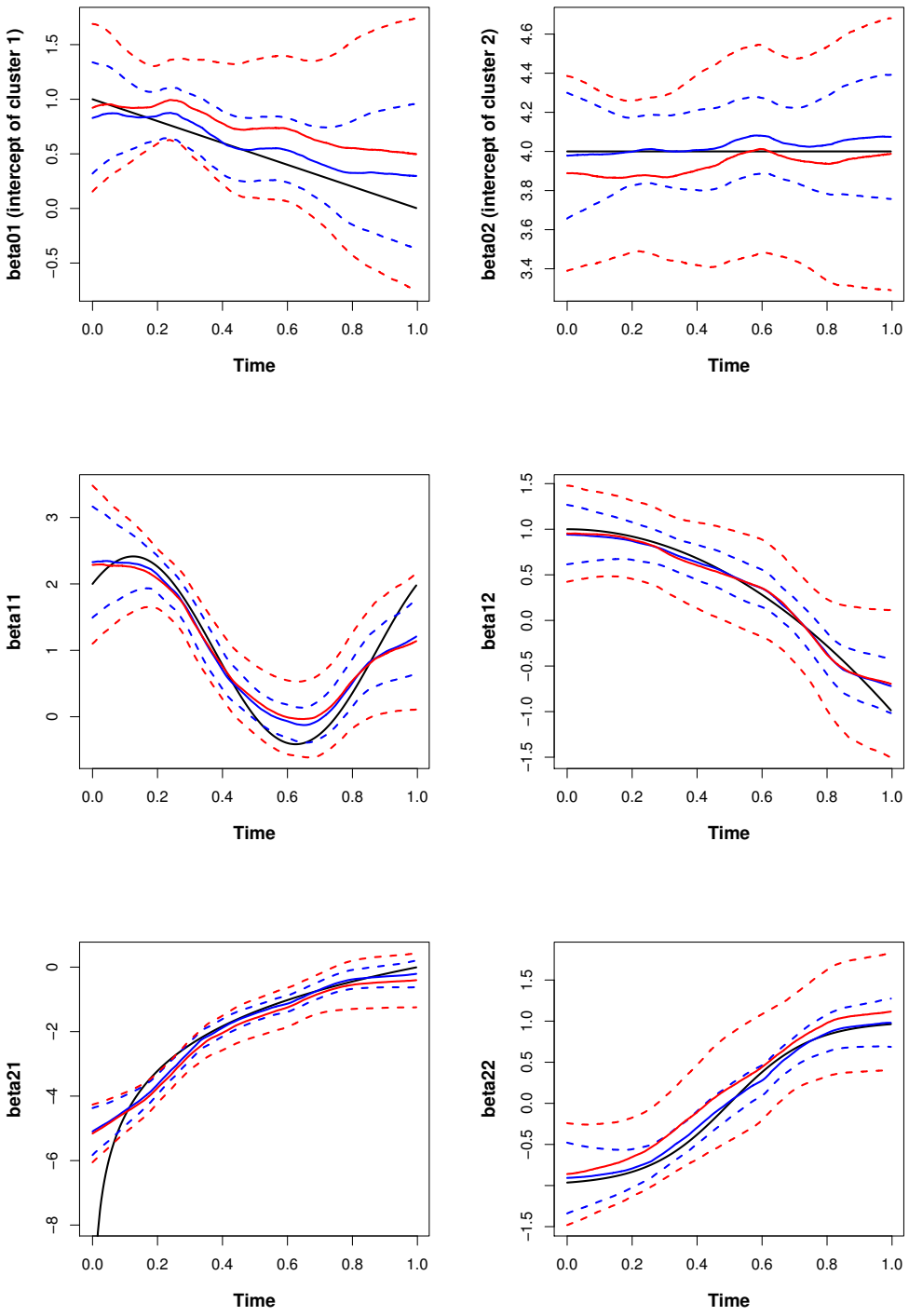

FIGURE 4.6 – The model coefficients and their 95% confidence interval estimated using 500 simulations with a sample size of N=500. Black curves represent the real coefficients. The solid curves in red represent the mean estimated time-varying coefficients obtained using the MVC model with no random effects while the dashed red curves are their 95% confidence intervals. The solid curves in blue are the mean estimated time-varying coefficients obtained using the model considering random effects. The dashed blue curves are their 95% confidence intervals. The first column contains the coefficients of the first mixture component ($\beta_{01}(t)$, $\beta_{11}(t)$ and $\beta_{21}(t)$ while the second column contains the coefficients obtained for the second component.

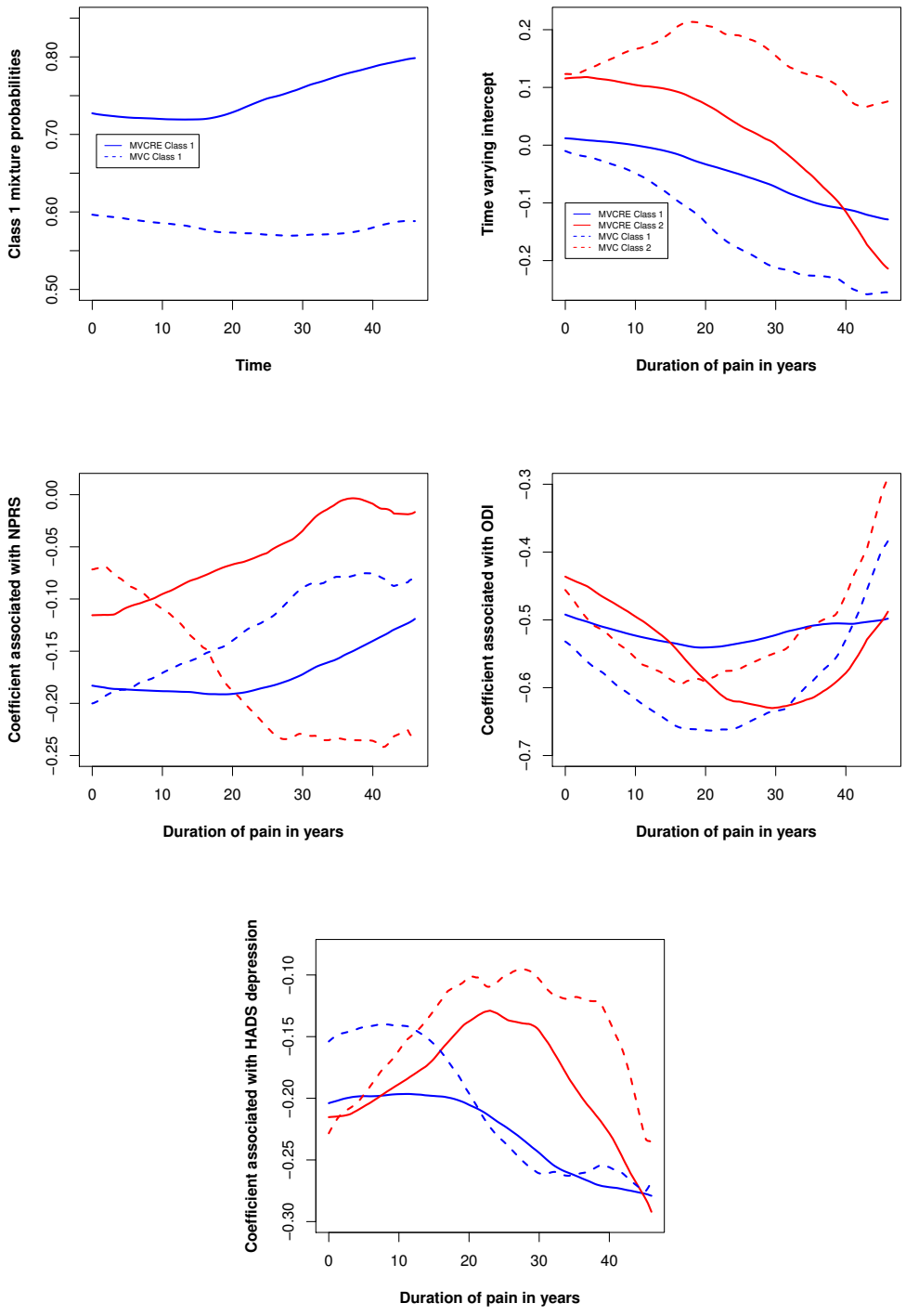

FIGURE 4.7 – Upper left: The solid blue curve represents class 1 mixture proportion as a function of time for the MVCRE model and the dashed lines represent the time-varying mixture proportions for the MVC model. Upper right and lower figures: The solid blue curves represent the time-varying coefficients obtained using the MVCRE model for class 1 observations while the dashed blue curves are the time-varying coefficient obtained using the MVC model for class1. The red curves represent the model coefficients obtained for class 2 patients using the MVCRE model. The dashed red lines represent the time-varying coefficients obtained using the MVC model for class 2 observations.

Chapitre 5

Mixture of longitudinal factor analysis for heterogeneous longitudinal multivariate data

Résumé du chapitre en français

Les données longitudinales multivariées jouent un rôle essentiel dans divers domaines de recherche, notamment les sciences médicales, sociales et comportementales, car elles permettent d'analyser l'évolution dans le temps de plusieurs indicateurs, mais aussi de déterminer comment ces changements sont influencés par d'autres variables. Ceci rends les techniques de modélisation dédiées aux données longitudinales multivariées très importantes. En général, pour étudier l'évolution de plusieurs indicateurs bruités parmi des patients échantillonnés dans une population homogène, il est important de se concentrer sur les tendances longitudinales parmi les variables latentes en utilisant une modélisation conjointe basée sur les structures de covariance entre les indicateurs observés. Cela permet aux chercheurs de tirer des conclusions plus générales et conjointes au lieu de conclure avec une analyse séparée pour chaque indicateur observé. Entre autre, le modèle d'analyse factorielle longitudinale peut accomplir cela. Ce modèle peut traiter à la fois les données non équilibrées et manquantes et intègre également la possibilité d'utiliser plus d'un facteur latent pour modéliser les indicateurs observés.
L'une des hypothèses de ce modèle est l'invariance de la structure factorielle entre les sous-groupes de sujets. Cette hypothèse restrictive suppose que la population étudiée est homogène en ce qui concerne la structure des facteurs latents étudiés. Cependant, il est possible que la structure des facteurs varie parmi des sous-groupes latents. Pour résoudre ce problème, certains auteurs ont proposé les modèles de mélange d'analyseurs factoriels. Ces modèles estiment différentes charges factorielles pour différentes sous-populations qui sont représentées par une variable de classe latente. A notre connaissance, aucune extension aux données longitudinales de ces modèles n'a été proposée. Dans ce chapitre, nous proposons une extension au cadre de l'analyse factorielle longitudinale où la non-invariance des groupes est prise en compte en utilisant un mélange d'analyseurs factoriels.

Le plan du chapitre est le suivant : nous commençons par définir le mélange du modèle d'analyse factorielle longitudinale et ses paramètres. Ensuite, nous proposons un algorithme EM pour estimer le modèle. Nous développons également un critère d'information bayésien pour identifier le nombre de composantes du mélange. Nous terminons en discutant de la comparabilité des facteurs latents obtenus entre les sujets de différents groupes latents.
Des simulations et des applications de données réelles ont été réalisées pour tester le modèle de mélange d'analyse factorielle longitudinale proposé dans ce chapitre et pour illustrer son applicabilité.

5.1 Introduction

Multivariate longitudinal data play a critical role in various research areas, including the medical, social and behavioral sciences because they not only allow the analysis of change over time of several indicators, but also identify how these changes are impacted by other variables. Several methods for multivariate longitudinal data analysis are available in the literature, and could be separated into two major categories: multivariate mixed effects models [4, 5] ; and factor analysis for longitudinal data including latent factor linear mixed models [6, 7], curve of factors models and factor of curves models [8], dynamic factor models and dynamic structural equation models (SEM) [9] and finally, latent process models [10]. Typically, these methods accept observed responses as the main interest variables. However, these response variables are generally just a noisy representation of latent underlying factors that cannot be observed or measured. An example is how depression is measured using a set of items containing measurement errors. Each one of these items represents a proportion of the information contained in the latent factor, which in this example is the severity of depressive symptoms.

Generally, in order to study the change in several noisy outcomes among patients sampled from a homogeneous population, it is important to focus on longitudinal trends among the latent variables using joint modeling based on the covariance structures of the observed outcomes. This allows researchers to draw more general and joint conclusions instead of conducting a separate analysis for each observed outcome.

Several of the previously cited models share the same measurement model, which is the equation representing the relationship between the measured/observed outcomes and the latent factors. On the other hand, the structural model representing the relationship between the latent factors differs between the models. For example, in the dynamic factor models, an autoregressive model is used to model the relationship between past and present factor structures, whereas in the latent factor linear mixed model, the latent factors are modeled as the outcome of a mixed effect model where polynomial functions of time and other covariates are considered as subject-specific effects. Another difference between these models is the consideration of a variant or invariant factor structure (factor loadings vary between groups of subjects or over time). Factor invariance is a classic assumption in factor analysis since latent constructs are supposed to be invariant in order to allow comparability, but in some domains, it is known that the structure of latent constructs changes due, for example, to cultural differences [11] or stage of a disorder [161].

The previously discussed methods are also different regarding the number of latent factors which can be included in the model and the handling of unbalanced data. The model proposed by An *et al.* [7] can handle both unbalanced and missing data and also can incorporate the possibility of using more than one latent factor, compared to the models proposed by Roy *et al.* [6] and Proust *et al.* [10]. One of the assumptions of the model proposed by An *et al.* [7] is invariance of the factors structure also called factorial time-invariance (*i.e.* factor loading matrix is assumed to be constant over time). Another more restrictive assumption is the invariance of factor structure among subgroups of subjects. This assumption supposes that the study population is homogeneous regarding the structure of the constructs being measured. In education and health sciences, it is possible that the structure of the factors vary among latent subpopulations. The impact of

measured variables on the underlying latent constructs can vary depending on other latent and observed characteristics. To address this issue, some authors have proposed mixture of factor analyzers [12]. These models estimate different factor loadings for different subpopulations that are represented by a latent class variable. To our knowledge, no extension to longitudinal data of these models has been proposed. In this chapter, we propose an extension to the longitudinal factor analysis framework where group non-invariance is taken into account using a mixture of factor analyzers.
The chapter layout is the following: we start by defining the mixture of longitudinal factor analysis model and its parameters. Then, we propose an EM algorithm to estimate the model. We also develop a Bayesian Information Criterion to identify the number of mixture components. We finish by discussing the comparability of the obtained latent factors between subjects from different latent groups. Simulations and real data applications were conducted to test the mixture of longitudinal factor analysis model proposed in this chapter and to illustrate its applicability.

5.2 Mixture of longitudinal factor analysis models (MLFA)

In this section we will give the formulation of our proposed model. We start by defining the longitudinal factor analysis model. Next, we generalize this longitudinal factor analysis model to heterogeneous data with varying constructs by including a mixture component. Throughout this chapter, we suppose that there are n subjects with J outcomes at t_{n_i} time points ($i = 1, \ldots, n$). The total number of observations will be noted $N = \sum_{i=1}^{n} n_i$. In the factor analysis framework, for subject i and time t, the outcome $(y_{ijt})_{j\in\{1,\ldots,J\}}$ can be described as a linear combination of K latent factors $(\eta_{ikt})_{k\in\{1,\ldots,K\}}$ ($K << J$) in the following manner:

$$y_{ijt} = \Lambda_j \eta_{it} + \epsilon_{ijt} \tag{5.1}$$

where Λ_j is the vector of factor loadings associated with the outcome $y_{.j.}$. The vector $\eta_{it} = (\eta_{i1t}, \ldots, \eta_{iKt})$ are the scores of the different latent factors for patient i at time t. We suppose that the variables in Y are standardized (to 0 mean and unit variance) in order to remove the scale differences across outcomes. The quantity ϵ_{ijt} is the error term also called unique factors, which follows the distribution $\mathcal{N}(0, \sigma_j^2)$. The proposed model assumes that factor loadings are constant over time. In other words, this assumption supposes time-invariance of our factors. This assumption ensures the comparability of the latent factors over time. Another assumption of the model is that $\mathbb{E}(\eta_{.k.}) = 0$ for all k. In other words, the latent factor means can vary across different time points but their global mean needs to be equal to 0.
The longitudinal evolution of the latent factors $\eta_{.k.}$ is then studied using the following multivariate mixed effects model:

$$\eta_{ikt} = X_{ikt}\beta_k + Z_{ikt}\xi_{ik} + \omega_{ikt} \tag{5.2}$$

where X_{ikt} and Z_{ikt} are respectively the fixed effects and random effects covariate matrices associated with the latent factor $\eta_{.kt}$, β_k are the fixed effects, and ξ_{ik} are the random effects

specific for subject i. The vector $\omega_{ikt} = (\omega_{ik1}, \ldots, \omega_{ikt_{n_i}})$ represents the error term for factor k and subject i. As can be observed in the model formulation, the covariate matrices X_k and Z_k can be different for different factors. This allows factors to be modeled separately and more flexibly.

5.2.1 Mixture of longitudinal factor analysis models

In this section we will introduce a model that accounts for latent group non-invariance/differences. In the classical factor analysis framework, group invariance is assumed in order to be able to conduct between-subject comparisons. This assumption is assured by supposing a unique factor structure (factor loadings matrix) for all the population. In our work, we assume that there exists latent groups between whom the factor structure is different. Our model is written as a mixture of the previously described longitudinal factor analysis models. The model can be written as follows:

$$y_{ijt} = \sum_{c=1}^{C} \mathbb{1}_{\{v_i=c\}}(\Lambda_{jc}\eta_{i.tc} + \epsilon_{ijtc}), \tag{5.3}$$

where v_i is a categorical latent variable representing the latent class of subject i. The variable v takes values in $\{1, \ldots, C\}$ associated with the probabilities $\{\pi_1, \ldots, \pi_C\}$ ($\sum_{c=1}^{C} \pi_c = 1$). The indicator function $\mathbb{1}_{\{v_i=c\}}$ is equal to 1 if subject i belongs to class c and 0 otherwise. From now on we denote $v_{ic} = \mathbb{1}_{\{v_i=c\}}$. Similar to the longitudinal factor analysis model discussed above, Λ_{jc} are the factor loadings associated with the outcome $y_{.jt}$ for cluster c. In the proposed model, the factor structure is allowed to be different among distinct heterogeneous subpopulations. The vector $\eta_{i.tc} = (\eta_{i1tc}, \ldots, \eta_{iKtc})$ contains the K latent factor scores for subject i in cluster c at time t. We assume that the number of latent factors is similar between classes.
We suppose that ϵ_{ijtc} follows the distribution $\mathcal{N}(0, \sigma^2_{jc})$ where σ^2_{jc} is the variance of the error terms associated with class c and the outcome j.
The latent factors for each class are modeled using the following mixed effects model:

$$\eta_{iktc} = X_{iktc}\beta_{kc} + Z_{iktc}\xi_{ikc} + \omega_{itc} \tag{5.4}$$

In the model represented by equation (5.4), β_{kc} are the fixed effects associated with the factor structure of the latent class c. The random effects associated with the latent class c are represented by the random variable ξ_{ikc} which are assumed to be normally distributed with mean 0 and variance covariance matrix Σ_{ξ_c}. The quantity ω_{itc} represents the error term of the structural model which follows a Gaussian distribution $\mathcal{N}(0, \Sigma_{\omega_c})$.
The proposed model has some advantages. Since we represent the J outcomes in y using a much smaller number K of latent factors in η, the estimation of the effects β is simpler than estimating these effects on the original J outcomes. Another advantage is that different latent factors and different classes can be modeled using different covariate matrices X_{kc} and Z_{kc}. This is useful since in practice, it can be of interest to study the effect of different variables on different latent factors or study different time-trajectories (linear, quadratic,...) for factors in different classes.
In equation 5.3, the mixture component allows different subpopulations to have different

latent factors that best reflect the relationship between their outcomes. However, this renders the model difficult to interpret when we try to compare the latent factors and their trajectories between groups. The inclusion of mixture parameters also generates a practical problem which is the identification of the optimal number of components C. These assumptions and problems will be discussed further.

5.3 Parameters estimation of the MLFA model with EM algorithm

In this section we discuss model estimation using the EM algorithm [79]. An advantage of using the EM algorithm for this model is that it is easy to implement, because closed form solutions are available for both E and M steps.
As a reminder, J is the number of outcomes y, K is the number of latent factors, n_i is the number of observations for subject i. Let p_c and q_c be the number of fixed and random effect covariates respectively for the structural model of class c factors $\eta_{.c}$. We start by defining the vectorized notations used in the estimation of our model:

$$y_{it} = (y_{i1t}, \ldots, y_{iJt}), \quad y_i = (y_{it_0}, \ldots, y_{it_{n_i}})$$

$$X_{itc} = \begin{pmatrix} X_{i1tc} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & X_{iKtc} \end{pmatrix}_{[K, p_c \times K]}, \quad X_{iktc} = (X_{iktc0}, \ldots, X_{iktcp_c})^T,$$

$$Z_{itc} = \begin{pmatrix} Z_{i1tc} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & Z_{iKtc} \end{pmatrix}_{[K, q_c \times K]}, \quad Z_{iktc} = (Z_{iktc0}, \ldots, Z_{iktcq_c})^T,$$

$$\eta_{ic} = (\eta_{itc_0}^T, \ldots, \eta_{itc_{n_i}}^T), \quad \eta_{itc}^T = (\eta_{i1tc}, \ldots, \eta_{iKtc}),$$

$$\Lambda_c = \begin{pmatrix} \Lambda_{1c}^T \\ \vdots \\ \Lambda_{Jc}^T \end{pmatrix}_{[J \times K]}$$

$$\xi_{ic} = (\xi_{ic1}^T, \ldots, \xi_{icK}^T), \quad \xi_{ick}^T = (\xi_{ick1}, \ldots, \xi_{ickq}),$$

$$\beta_c = (\beta_{c1}^T, \ldots, \beta_{cK}^T), \quad \beta_{ck}^T = (\beta_{ck1}, \ldots, \beta_{ckp}),$$

$$\pi = \begin{pmatrix} \pi_1 \\ \vdots \\ \pi_C \end{pmatrix}.$$

The log-likelihood of the observed data $\{y, X, Z, \eta, \xi\}$ is the following:

$$\log L(\theta) = \sum_{i=1}^{n} \log(\sum_{c=1}^{C} \pi_c \phi(y_i, X_{ic}, Z_{ic}, \eta_{ic}, \xi_{ic}, \theta_c)). \tag{5.5}$$

The maximization of this log-likelihood directly is difficult due to the presence unobserved variables. Therefore, we use the EM algorithm to maximize this log-likelihood indirectly using the following expected complete data log-likelihood:

$$\mathbb{E}(\log L_{comp}(\theta)|y_i) = \sum_{i=1}^{n}\sum_{c=1}^{C} \mathbb{E}(v_{ic}\log(\pi_c) + v_{ic}\log(\phi(y_i, X_i, Z_{ic}, \eta_{ic}, \xi_{ic}, \theta_c))) + constant \tag{5.6}$$

We have the following:

$$\begin{aligned}
\log(\phi(y_i, X_i, Z_{ic}, \eta_{ic}, \xi_{ic}, \theta_c)) &= \log(\phi(y_i|\eta_{ic}, \theta_c)) + \log(\phi(\eta_{ic}|\xi_{ic}, \theta_c)) + \log(\phi(\xi_{ic}, \theta_c)) \\
&= \sum_{j=1}^{J}\{-\frac{t_{n_i}}{2}\log\sigma_{jc}^2 - \frac{1}{2\sigma_{jc}^2}(y_{ij} - \eta_{ic}^T\Lambda_{jc})^T(y_{ij} - \eta_{ic}^T\Lambda_{jc})\} \\
&- \frac{1}{2}\sum_{t=1}^{t_{n_i}}(\log|\Sigma_\omega| + (\eta_{itc} - Z_{itc}\xi_{ic} - X_{itc}\beta_c)^T\Sigma_\omega^{-1}(\eta_{itc} - Z_{itc}\xi_{ic} - X_{itc}\beta_c)) \\
&- (\frac{1}{2}\log|\Sigma_{\xi_c}| + \frac{1}{2}\xi_{ic}^T\Sigma_{\xi_c}^{-1}\xi_{ic}).
\end{aligned}$$

In order to be able to maximize the expected log-likelihood, the following conditional expectations of the interactions between the hidden variables (sufficient statistics) need to be calculated: $\mathbb{E}(v_{ic}|y_i)$, $\mathbb{E}(\eta_{itc}|y_i)$, $\mathbb{E}(v_{ic}\eta_{itc}|y_i)$, $\mathbb{E}(v_{ic}\eta_{itc}\eta_{itc}^T|y_i)$, $\mathbb{E}(\eta_{itc}\eta_{itc}^T|y_i)$, $\mathbb{E}(\xi_{ic}|y_i)$, $\mathbb{E}(v_{ic}\xi_{ic}|y_i)$, $\mathbb{E}(v_{ic}\xi_{ic}\xi_{ic}^T|y_i)$, $\mathbb{E}(\xi_{ic}\xi_{ic}^T|y_i)$ and $\mathbb{E}(\eta_{itc}\xi_{ic}^T|y_i)$. We can easily verify that:

$$\begin{aligned}
\mathbb{E}(v_{ic}\xi_{ic}|y_i) &= \mathbb{E}(v_{ic}|y_i)\,\mathbb{E}(\xi_{ic}|v_{ic}, y_i), \\
\mathbb{E}(v_{ic}\eta_{itc}|y_i) &= \mathbb{E}(v_{ic}|y_i)\,\mathbb{E}(\eta_{itc}|v_{ic}, y_i), \\
\mathbb{E}(v_{ic}\xi_{ic}\xi_{ic}^T|y_i) &= \mathbb{E}(v_{ic}|y_i)\,\mathbb{E}(\xi_{ic}\xi_{ic}^T|v_{ic}, y_i), \\
\mathbb{E}(v_{ic}\eta_{itc}\eta_{itc}^T|y_i) &= \mathbb{E}(v_{ic}|y_i)\,\mathbb{E}(\eta_{itc}\eta_{itc}^T|v_{ic}, y_i).
\end{aligned}$$

The maximization of the complete log-likelihood given in equation (5.5) gives the following estimation equations for the measurement model (in equation (5.3)):

$$\Lambda_c = ((\sum_{i=1}^{n}\sum_{t=1}^{t_{n_i}} \mathbb{E}(v_{ic}|y_i)\,\mathbb{E}(\eta_{itc}\eta_{itc}^T|v_{ic}, y_i)^{-1}(\sum_{i=1}^{n}\sum_{t=1}^{t_{n_i}} \mathbb{E}(v_{ic}|y_i)\,\mathbb{E}(\eta_{itc}|v_{ic}, y_i)y_{it}^T))^T, \tag{5.7}$$

$$\begin{aligned}
\sigma_{jc} =& \sqrt{\frac{1}{\sum_{i=1}^{n} n_i\,\mathbb{E}(v_{ic}|y_i)} diag(\sum_{i=1}^{n}\sum_{t=1}^{t_{n_i}} \mathbb{E}(v_{ic}|y_i)\,\mathbb{E}((y_{ic} - \hat{\Lambda}_c\eta_{itc})(y_{ic} - \hat{\Lambda}_c\eta_{itc})^T|y_i, v_{ic}))} \\
=& (\frac{1}{\sum_{i=1}^{n} n_i\,\mathbb{E}(v_{ic}|y_i)} diag(\sum_{i=1}^{n}\sum_{t=1}^{t_{n_i}} \mathbb{E}(v_{ic}|y_i)y_{ij}y_{it}^T - \mathbb{E}(v_{ic}|y_i)y_{it}\,\mathbb{E}(\eta_{itc}|v_{ic}, y_i)^T\hat{\Lambda}_c^T \\
&- \mathbb{E}(v_{ic}|y_i)\hat{\Lambda}_c\,\mathbb{E}(\eta_{itc}|v_{ic}, y_i)y_{it}^T + \mathbb{E}(v_{ic}|y_i)\hat{\Lambda}_c\,\mathbb{E}(\eta_{itc}\eta_{itc}^T|y_i, v_{ic})\hat{\Lambda}_c^T))^{1/2}.
\end{aligned} \tag{5.8}$$

For the structural model given in equation (5.4), we have the following estimation equations:

$$\Sigma_{\xi_c} = \frac{1}{\sum_{i=1}^{n} \mathbb{E}(v_{ic}|y_i)} \sum_{i=1}^{n} \mathbb{E}(v_{ic}|y_i)\, \mathbb{E}(\xi_{ic}\xi_{ic}^T|y_i, v_{ic}). \tag{5.9}$$

The fixed effects can be estimated by:

$$\beta_c = (\sum_{i=1}^{n}\sum_{t=1}^{t_{n_i}} \mathbb{E}(v_{ic}|y_i) X_{itc}^T \Sigma_{\omega_c}^{-1} X_{itc})^{-1} (\sum_{i=1}^{n}\sum_{t=1}^{t_{n_i}} \mathbb{E}(v_{ic}|y_i) X_{itc}^T \Sigma_{\omega_c}^{-1} (\mathbb{E}(\eta_{itc}|v_{ic}, y_i) - Z_{itc}\, \mathbb{E}(\xi_{ic}|v_{ic}, y_i))), \tag{5.10}$$

and

$$\begin{aligned}
\Sigma_{\omega_c} &= \frac{1}{\sum_{i=1}^{n} n_i\, \mathbb{E}(v_{ic}|y_i)} \sum_{i=1}^{n}\sum_{t=1}^{t_{n_i}} \mathbb{E}(v_{ic}|y_i)\, \mathbb{E}((\eta_{itc} - X_{itc}\beta_c - Z_{itc}\xi_{ic})(\eta_{itc} - X_{itc}\beta_c - Z_{itc}\xi_{ic})^T|y_i, v_{ic}) \\
&= \frac{1}{\sum_{i=1}^{n} n_i\, \mathbb{E}(v_{ic}|y_i)} \sum_{i=1}^{n}\sum_{t=1}^{t_{n_i}} \mathbb{E}(v_{ic}|y_i)\, \mathbb{E}(\eta_{itc}\eta_{itc}^T|y_i, v_{ic}) - \mathbb{E}(v_{ic}|y_i) X_{itc}\beta_c\, \mathbb{E}(\eta_{itc}^T|y_i, v_{ic}) \\
&- \mathbb{E}(v_{ic}|y_i) Z_{itc}\, \mathbb{E}(\xi_{ic}\eta_{itc}^T|y_i, v_{ic}) - \mathbb{E}(v_{ic}|y_i)\, \mathbb{E}(\eta_{itc}|y_i, v_{ic})\beta_c^T X_{itc}^T \\
&- \mathbb{E}(v_{ic}|y_i)\, \mathbb{E}(\eta_{itc}\xi_{ic}^T|y_i, v_{ic}) Z_{itc}^T + \mathbb{E}(v_{ic}|y_i) X_{itc}\beta_c\, \mathbb{E}(\xi_{ic}^T|y_i, v_{ic}) Z_{itc}^T \\
&+ \mathbb{E}(v_{ic}|y_i) Z_{itc}\, \mathbb{E}(\xi_{ic}|y_i, v_{ic})\beta_c^T X_{itc}^T + \mathbb{E}(v_{ic}|y_i) X_{itc}\beta_c\beta_c^T X_{itc}^T \\
&+ \mathbb{E}(v_{ic}|y_i) Z_{itc}\, \mathbb{E}(\xi_{ic}\xi_{ic}^T|y_i, v_{ic}) Z_{itc}^T,
\end{aligned} \tag{5.11}$$

and the probability class membership can be obtained as:

$$\pi_c = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}(v_{ic}|y_i). \tag{5.12}$$

We summarize the proposed EM procedure discussed above in Algorithm 3. The conditional expectations of the sufficient statistics are given in the appendix of this chapter.

5.4 Number of mixture components

Similar to the mixture model proposed in chapter 4, we propose a BIC criterion to identify the optimal number C of mixture components. the BIC of the model could be calculated as:

$$BIC_{MLFA} = -2\log L(\theta) + (C(\#parameters) + C - 1)\log(N) \tag{5.13}$$

where $\#parameters = (JK + J + \frac{K(K+1)}{2} + K(p+1) + \frac{(q+1)K((q+1)K+1)}{2})$ is the number of parameters in the model. JK is the number of free parameters in the factor loadings matrix. J is the number of residual variances in σ_c. $\frac{K(K+1)}{2}$ is the number of estimated parameters in covariance matrix Σ_{ω_c}. $p+1$ is the number of fixed effect coefficients (p covariates and intercept) and $\frac{(q+1)K((q+1)K+1)}{2}$ is the number of parameters in the covariance matrix of the random effects. The number of parameters of the mixture proportions is $C-1$.
We will show further in our simulation studies that this BIC criterion allows us to identify the optimal number of mixture components. We do this by fitting the model with different numbers of components and select the model yielding the smallest BIC_{MLFA}.

Algorithm 3 The EM algorithm for estimating the MLFA model

— **Input**: Number of components C and number of latent factors K, subject identifiers (noted by $i = 1, \ldots, n$), the J outcomes (y_{ijt} $i = 1, \ldots, n, j = 1, \ldots, J, t = 1, \ldots, n_i$, fixed effects covariates (X_{ikt}) $i = 1, \ldots, n, k = 1, \ldots, K, t = 1, \ldots, n_i$ and random effects covariates (Z_{ikt}) $i = 1, \ldots, n, k = 1, \ldots, K, t = 1, \ldots, n_i$, and maximum number of iterations max_iter.
— Initialize the parameters $\theta_c = (\pi, \Lambda, \sigma, \beta, \Sigma_\xi, \Sigma_\omega)_c$ for $c = 1, \ldots, C$.
— *E-step*: calculate the conditional expectations $\mathbb{E}(v_{ic}|y_i)$, $\mathbb{E}(\eta_{itc}|y_i, v_{ic})$, $\mathbb{E}(\eta_{itc}\eta_{itc}^T|y_i, v_{ic})$, $\mathbb{E}(\xi_{ic}|y_i, v_{ic})$, $\mathbb{E}(\xi_{ic}\xi_{ic}^T|y_i, v_{ic})$ and $\mathbb{E}(\eta_{itc}\xi_{ic}^T|y_i, v_{ic})$.
— *M-step*: insert the quantities calculated in the *E-step* in the equations (5.7), (5.8), (5.9) and (5.11) to get the maximization solutions of the expected log-likelihood $\mathbb{E}(\log L(\theta)|y_i)$.
— Repeat the *E-step* and *M-step* until convergence or the maximum number of iterations max_iter is reached.
— **Output**: The estimations of the parameters in θ: $\hat{\Lambda}_c$, $\hat{\sigma_c}$, $\hat{\Sigma}_{\omega_c}$, $\hat{\beta}_c$, $\hat{\Sigma}_{\xi_c}$, $\hat{\pi}_c(t)$ for $c = 1, \ldots, C$.

5.5 Simulations study

We used simulated data to evaluate the computational and practical properties of the proposed model. We also evaluate the performance of the BIC criterion proposed in this chapter at identifying the number of mixture components. We repeat $S = 100$ simulations. We evaluate the accuracy of the parameters estimates using Mean Absolute Error ($MAE = \frac{1}{S}\sum_{s=1}^{S}|p - \hat{p}_s|$, where p is the true parameter and $\hat{p}_s$ is the estimated parameter at simulation s). We take a two-component mixture of longitudinal factor analysis models. Three different simulation studies were conducted. The first simulation study was conducted on balanced data where factor loadings are the same across items in a given factor and the observations are obtained at the same time points for different subjects. This study includes 500 subjects with 5 observations each. Two latent factors were considered in the simulations. These two factors were measured using 10 items. A mixed effects model with random intercept and slopes was used to simulate the relationship between the latent factors and the explanatory variables. We also assumed a simple structure to model the relationship between these latent factors and the items, which suggests that each item loads on only one of the latent factors. This assumption is generally applied in practice in order to simplify the interpretability of the factor structure. This is achieved by fixing the last 5 elements of the first column and the first 5 elements of the second column of the factor loading matrix at 0. To fix the scale of latent factors, $\lambda_{1,1,c}$ and $\lambda_{6,2,c}$ are fixed to 1 for $c = 1, 2$. The following parameters were considered in the simulations:

$$C = 2, \qquad t \in \{1, 2, 3, 4, 5\},$$

$$\pi_1 = 0.6,\ \pi_2 = 1 - \pi_1,\ \Lambda_1 = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \Lambda_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix},$$

$$\sigma^2_{j1} = 1 \text{ for all } j \in \{1, \ldots, 10\} \text{ and } \sigma^2_{j2} = 0.75^2 \text{ for all } j \in \{1, \ldots, 10\},$$
$$X \sim \mathcal{N}((0,0)^T,\ I_2), \qquad Z = X,$$
$$\beta_{111} = -1, \qquad \beta_{112} = 1,$$
$$\beta_{121} = 1, \qquad \beta_{122} = -1,$$
$$\beta_{211} = 1, \qquad \beta_{212} = -1,$$
$$\beta_{221} = 0, \qquad \beta_{222} = -2,$$
$$\xi_{ickp} = a_{ickp} \text{ for } i \text{ in } \{1, \ldots, 500\} \text{ and } k \text{ in } \{1, 2\},$$
$$\text{where } (a_{ic11}, a_{ic12}, a_{ic21}, a_{ic22})^T \sim N((0,0,0,0)^T, \Sigma_\xi) \text{ for } c \text{ in } \{1, 2\},$$
$$\omega_{ic} \sim \mathcal{N}((0,0)^T, \Sigma_{\omega_c}),$$

where $\Sigma_\xi = \begin{pmatrix} 1 & 0.5^2 & 0.5^2 & 0.5^2 \\ 0.5^2 & 1 & 0.5^2 & 0.5^2 \\ 0.5^2 & 0.5^2 & 1 & 0.5^2 \\ 0.5^2 & 0.5^2 & 0.5^2 & 1 \end{pmatrix}$ and $\Sigma_{\omega_1} = \begin{pmatrix} 0.5^2 & 0.2^2 \\ 0.2^2 & 0.5^2 \end{pmatrix}$ for $c = 1$ and $\Sigma_{\omega_2} = \begin{pmatrix} 0.75^2 & 0.5^2 \\ 0.5^2 & 0.75^2 \end{pmatrix}$ for $c = 2$. We considered both subject-specific intercept and slopes.
The final simulated model can then be written as:

$$y_{ijt} = \sum_{c=1}^{C} \mathbb{1}_{\{v_i = c\}} (\Lambda_{jc} \eta_{itc} + \epsilon_{ijtc}),$$

where v_i is the categorical variable representing the class to which subject i belongs (with probability 0.6 for $c = 1$ and probability 0.4 for $c = 2$). For $c = 1$ and for the first factor we have:

$$\eta_{it11} = -1 * X_{i1} + 1 * X_{i1} + Z_{i1}\xi_{i111} + Z_{i2}\xi_{i112} + \omega_{i11t},$$

for $c = 1$ and for the second factor we have:

$$\eta_{it12} = 1 * X_{i1} - 1 * X_{i1} + Z_{i1}\xi_{i121} + Z_{i2}\xi_{i122} + \omega_{i12t},$$

for $c = 2$ and for the first factor we have:

$$\eta_{it21} = 1 * X_{i1} - 1 * X_{i1} + Z_{i1}\xi_{i211} + Z_{i2}\xi_{i212} + \omega_{i21t},$$

for $c = 2$ and for the second factor we have:

$$\eta_{it22} = 0 * X_{i1} - 2 * X_{i1} + Z_{i1}\xi_{i221} + Z_{i2}\xi_{i222} + \omega_{i22t}.$$

A model graphical representation of the simulated data is given in figure 5.1. For these simulations, the initial values for the factor loadings were generated randomly from a uniform distribution in $[0, 1]$. The fixed effects were initiated randomly following a standard normal distribution. The variance and covariance matrices were initiated using identity matrices for all components. Finally, the specific factors variance was set to 1.
In the second simulation we used the same setting as the first one except that data were simulated in an unbalanced manner (*i.e.* observations were measured at different time points for different patients and the number of observations differed between patients) and the factor loadings in the same factor might differ between items. The following factor loadings matrices were used for this simulation:

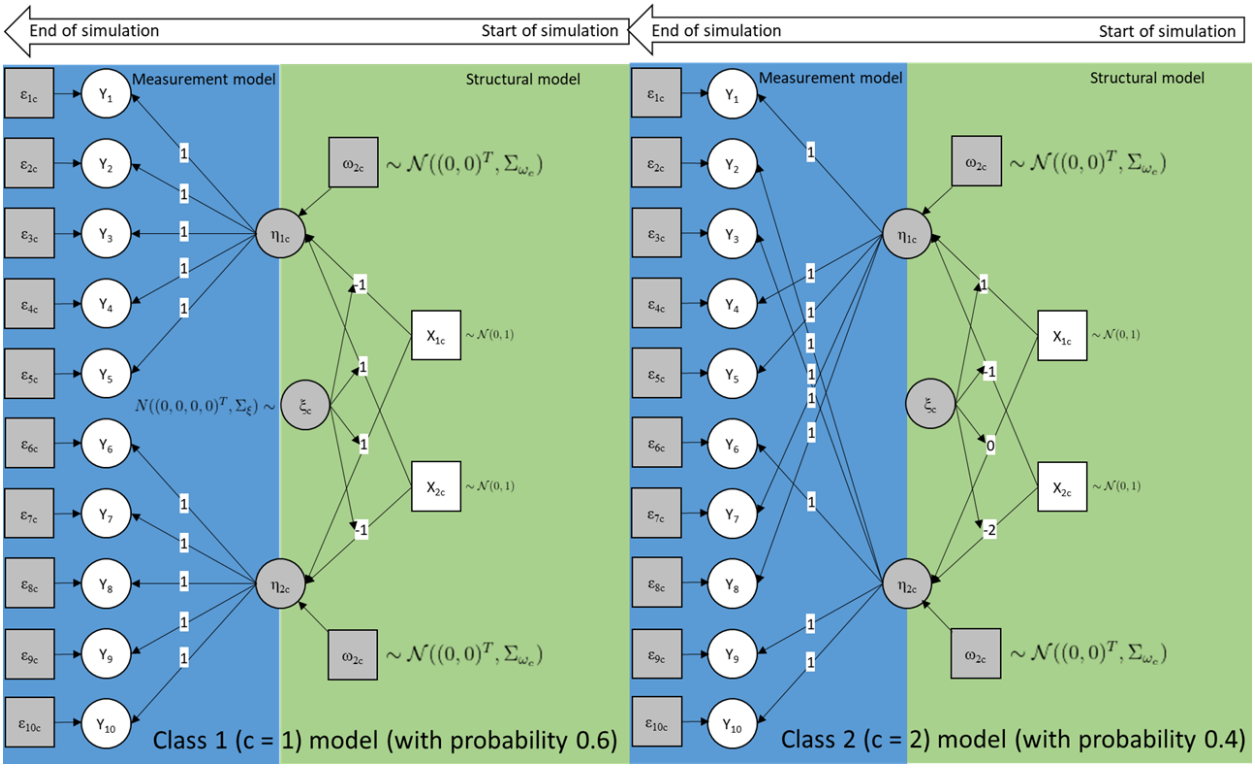

FIGURE 5.1 – A graphical model representation of the simulated data. This representation includes the measurement model (in the blue square) and structural model (in the yellow square). The quantities in a white circle represent the observed/measured variables. The quantities in a grey circle represent the latent variables (latent factors and random effects). The quantities in a white square represent the structural mixed effects model covariates. Grey squares represent the error terms. An arrow represents either a regression or a loading.

$$\Lambda_1 = \begin{pmatrix} 1 & 0 \\ 0.8 & 0.2 \\ 0.7 & 0.3 \\ 0.9 & 0.1 \\ 0.6 & 0.4 \\ 0 & 1 \\ 0.7 & 0.3 \\ 0.8 & 0.2 \\ 0.9 & 0.1 \\ 0.2 & 0.8 \end{pmatrix}, \qquad \Lambda_2 = \begin{pmatrix} 1 & 0 \\ 0.3 & 0.7 \\ 0.2 & 0.8 \\ 0.6 & 0.4 \\ 0.7 & 0.3 \\ 1 & 0 \\ 0.9 & 0.1 \\ 0.7 & 0.3 \\ 0.8 & 0.2 \\ 0.1 & 0.9 \end{pmatrix}.$$

For this simulation, time points were simulated using a uniform distribution in $[0, 1]$. Each time point was assigned to a subject using a discrete uniform distribution.

In the third simulation, we conducted 100 simulations with balanced data using only the first component parameters (*i.e.* assuming a homogeneous sample). We then modeled this 100 1-component and 2-component simulations using 1, 2, 3 and 4-component MLFA. The four models were compared using the proposed BIC_{MLFA} criterion to assess its utility in the selection of the number of components.

Results of the first simulation are presented in Table 5.1. The MAE of the factor loadings ranged from 0.013 to 0.018 for both classes. The model was able to separate the subjects in the two classes. The mean posterior probability of subjects in the first component was 0.84. The mean posterior probability of subjects in the second component was 0.106. The variance/covariance components were also estimated adequately with a MAE ranging from

0.014 to 0.089. The fixed effect parameters MAE ranged from 0.040 to 0.127. Therefore, the proposed EM algorithm is able to adequately estimate the model parameters. Regarding the mixed effect model, we also obtained good parameters estimates. The MAEs obtained for the unbalanced data simulations (5.2) were similar to those obtained for the balanced data simulation.
Regarding the evaluation of the BIC criterion, we found that the proposed BIC_{MLFA} can identify the optimal number of mixture components. The number of mixture components was correctly identified for 100% of the simulations for the model with a single component. For the model with two components, the number of mixture components was correctly identified for 75% of the simulations.

5.6 Real data example

In this section, we propose to apply our model to the PREDIBACK study data discussed in the previous chapters. As a reminder, the PREDIBACK study comprises of clinical, sociodemographic and cognitive-behavioural variables of 200 patients with chronic pain after spinal surgery, evaluated each 3 months over a 1-year follow-up.
In chapter 3 ([41]), we showed that quality of life of different clusters is impacted differently by pain intensity and by functional disability. This suggests that the structure/model relating different pain dimensions might change between clusters of patients with specific characteristics. The majority of evaluation tools used in the pain literature are developed and validated on chronic pain populations which do not share the same goals regarding pain treatments. For example, the Oswestry Disability Index questionnaire which includes items such as "walking" and "standing", is used to evaluate chronic pain patients including patients with amputed lower limbs. These patients score lower on these items even when adequate pain relief is achieved. Therefore, weighting such items lower when evaluating the disability level of these patients would give a better representation of the true evolution of their disability following a therapy [162]. In a paper by Rigoard et al. [28], we have proposed a factor analysis model using 25 items evaluating pain intensity, disability, anxiety/depression, quality of life and pain surface. The factor loadings were supposed to be invariant across all groups of patients (metric invariance). This assumption is hard to verify since the factor loadings might vary between groups which are not observed and not clearly identified. Although the score derived from the previous paper showed a good performance at evaluating patients well being, the invariance assumptions were not theoretically founded. Therefore, it is important to identify these latent groups for whom the structure representing the relationships between pain dimensions differ.
In this application, we propose a mixture of latent factors mixed effect model to assess the evolution of multiple outcomes in a heterogeneous population. We consider the situation where these observed longitudinal outcomes estimate several underlying longitudinal latent outcomes with structures that vary between individuals based on their biopsychosocial characteristics. The latent outcomes were modeled using the following observed outcomes: pain intensity measured using the NPRS, functional capacity assessed using the 10 items of the ODI questionnaire, depression and anxiety assessed using the 14 items of the HADS questionnaire, quality of life measured using the 5 items of the EQ-5D and the pain surface measured using a tactile interface where the patient could draw their pain using 4 different

TABLE 5.1 – Results of the first simulation study with balanced data ($N = 2500$). This table shows the mean absolute error (MAE) obtained from the 100 simulations and its standard deviation (SD) for each simulated parameter of the model and for each of the two latent classes. We can observe that the mean absolute errors were relatively small for all the parameters.

Data type	Component 1 parameters			Component 2 parameters		
	Parameter	True value	MAE (SD of AE)	Parameter	True value	MAE (SD of AE)
Balanced	π_1	0.6	0.018 (0.014)	π_2	-	-
	$\lambda_{2,1,1}$	1	0.017 (0.012)	$\lambda_{2,2,2}$	1	0.015 (0.012)
	$\lambda_{3,1,1}$	1	0.015 (0.013)	$\lambda_{3,2,2}$	1	0.013 (0.010)
	$\lambda_{4,1,1}$	1	0.018 (0.014)	$\lambda_{4,1,2}$	1	0.015 (0.012)
	$\lambda_{5,1,1}$	1	0.017 (0.014)	$\lambda_{5,1,2}$	1	0.017 (0.013)
	$\lambda_{7,2,1}$	1	0.016 (0.013)	$\lambda_{7,1,2}$	1	0.015 (0.012)
	$\lambda_{8,2,1}$	1	0.017 (0.013)	$\lambda_{8,1,2}$	1	0.016 (0.014)
	$\lambda_{9,2,1}$	1	0.016 (0.011)	$\lambda_{9,2,2}$	1	0.014 (0.010)
	$\lambda_{10,2,1}$	1	0.017 (0.012)	$\lambda_{10,2,2}$	1	0.013 (0.010)
	$\sigma_{1,1}$	1	0.021 (0.015)	$\sigma_{1,2}$	0.75	0.031 (0.024)
	$\sigma_{2,1}$	1	0.023 (0.016)	$\sigma_{2,2}$	0.75	0.036 (0.030)
	$\sigma_{3,1}$	1	0.020 (0.015)	$\sigma_{3,2}$	0.75	0.033 (0.027)
	$\sigma_{4,1}$	1	0.018 (0.014)	$\sigma_{4,2}$	0.75	0.035 (0.026)
	$\sigma_{5,1}$	1	0.019 (0.014)	$\sigma_{5,2}$	0.75	0.036 (0.028)
	$\sigma_{6,1}$	1	0.022 (0.015)	$\sigma_{6,2}$	0.75	0.033 (0.022)
	$\sigma_{7,1}$	1	0.022 (0.014)	$\sigma_{7,2}$	0.75	0.034 (0.024)
	$\sigma_{8,1}$	1	0.019 (0.016)	$\sigma_{8,2}$	0.75	0.033 (0.023)
	$\sigma_{9,1}$	1	0.019 (0.016)	$\sigma_{9,2}$	0.75	0.035 (0.028)
	$\sigma_{10,1}$	1	0.019 (0.013)	$\sigma_{10,2}$	0.75	0.034 (0.025)
	$\beta_{1,1,1}$	-1	0.127 (0.039)	$\beta_{2,1,1}$	1	0.042 (0.032)
	$\beta_{1,1,2}$	1	0.040 (0.026)	$\beta_{2,1,2}$	-1	0.063 (0.040)
	$\beta_{1,2,1}$	1	0.049 (0.028)	$\beta_{2,2,1}$	-1	0.059 (0.039)
	$\beta_{1,2,2}$	0	0.101 (0.038)	$\beta_{2,2,2}$	-2	0.109 (0.050)
	$\Sigma_{\omega_{1,1,1}}$	0.25	0.019 (0.014)	$\Sigma_{\omega_{1,1,2}}$	0.56	0.028 (0.021)
	$\Sigma_{\omega_{1,2,1}}$	0.04	0.014 (0.011)	$\Sigma_{\omega_{1,2,2}}$	0.25	0.017 (0.013)
	$\Sigma_{\omega_{2,2,1}}$	0.25	0.020 (0.016)	$\Sigma_{\omega_{2,2,2}}$	0.56	0.026 (0.021)
	$\Sigma_{\xi_{1,1,1}}$	1	0.067 (0.051)	$\Sigma_{\xi_{1,1,2}}$	1	0.091 (0.066)
	$\Sigma_{\xi_{1,2,1}}$	0.25	0.058 (0.046)	$\Sigma_{\xi_{1,2,2}}$	0.25	0.048 (0.039)
	$\Sigma_{\xi_{1,3,1}}$	0.25	0.046 (0.039)	$\Sigma_{\xi_{1,3,2}}$	0.25	0.051 (0.036)
	$\Sigma_{\xi_{1,4,1}}$	0.25	0.052 (0.034)	$\Sigma_{\xi_{1,4,2}}$	0.25	0.056 (0.037)
	$\Sigma_{\xi_{2,2,1}}$	0.25	0.089 (0.068)	$\Sigma_{\xi_{2,2,2}}$	0.25	0.072 (0.051)
	$\Sigma_{\xi_{2,3,1}}$	0.25	0.050 (0.041)	$\Sigma_{\xi_{2,3,2}}$	0.25	0.054 (0.038)
	$\Sigma_{\xi_{2,4,1}}$	0.25	0.046 (0.029)	$\Sigma_{\xi_{2,4,2}}$	0.25	0.048 (0.035)
	$\Sigma_{\xi_{3,3,1}}$	1	0.063 (0.050)	$\Sigma_{\xi_{3,3,2}}$	1	0.074 (0.057)
	$\Sigma_{\xi_{3,4,1}}$	0.25	0.050 (0.033)	$\Sigma_{\xi_{3,4,2}}$	0.25	0.050 (0.038)
	$\Sigma_{\xi_{4,4,1}}$	1	0.073 (0.055)	$\Sigma_{\xi_{4,4,2}}$	1	0.072 (0.058)

TABLE 5.2 – Results of the second simulation study with unbalanced data ($N = 2500$). This table shows the mean absolute error (MAE) obtained from the 100 simulations and its standard deviation (SD) for each simulated parameter of the model and for each of the two latent classes. The relatively small MAEs were also maintained for unbalanced data.

Data type	Component 1 parameters			Component 2 parameters		
	Parameter	True value	MAE (SD of AE)	Parameter	True value	MAE (SD of AE)
Unbalanced	π_1	0.6	0.032 (0.006)	π_2	-	-
	$\lambda_{2,1,1}$	0.8	0.014 (0.010)	$\lambda_{2,1,2}$	0.3	0.013 (0.010)
	$\lambda_{3,1,1}$	0.7	0.014 (0.011)	$\lambda_{3,1,2}$	0.2	0.012 (0.010)
	$\lambda_{4,1,1}$	0.9	0.016 (0.013)	$\lambda_{4,1,2}$	0.6	0.013 (0.009)
	$\lambda_{5,1,1}$	0.6	0.014 (0.010)	$\lambda_{5,1,2}$	0.7	0.014 (0.011)
	$\lambda_{7,1,1}$	0.7	0.013 (0.010)	$\lambda_{7,1,2}$	0.9	0.015 (0.011)
	$\lambda_{8,1,1}$	0.8	0.016 (0.010)	$\lambda_{8,1,2}$	0.7	0.013 (0.011)
	$\lambda_{9,1,1}$	0.9	0.015 (0.011)	$\lambda_{9,1,2}$	0.8	0.012 (0.010)
	$\lambda_{10,1,1}$	0.2	0.017 (0.012)	$\lambda_{10,1,2}$	0.1	0.013 (0.009)
	$\lambda_{2,2,1}$	0.2	0.022 (0.016)	$\lambda_{2,2,2}$	0.7	0.012 (0.010)
	$\lambda_{3,2,1}$	0.3	0.022 (0.016)	$\lambda_{3,2,2}$	0.8	0.012 (0.008)
	$\lambda_{4,2,1}$	0.1	0.019 (0.013)	$\lambda_{4,2,2}$	0.4	0.012 (0.009)
	$\lambda_{5,2,1}$	0.4	0.015 (0.011)	$\lambda_{5,2,2}$	0.3	0.010 (0.007)
	$\lambda_{7,2,1}$	0.3	0.016 (0.012)	$\lambda_{7,1,2}$	0.1	0.011 (0.009)
	$\lambda_{8,2,1}$	0.2	0.018 (0.015)	$\lambda_{8,2,2}$	0.3	0.012 (0.008)
	$\lambda_{9,2,1}$	0.1	0.018 (0.015)	$\lambda_{9,2,2}$	0.2	0.011 (0.009)
	$\lambda_{10,2,1}$	0.8	0.016 (0.011)	$\lambda_{10,2,2}$	0.9	0.012 (0.008)
	$\sigma_{1,1}$	1	0.023 (0.017)	$\sigma_{1,2}$	0.75	0.046 (0.034)
	$\sigma_{2,1}$	1	0.023 (0.018)	$\sigma_{2,2}$	0.75	0.044 (0.036)
	$\sigma_{3,1}$	1	0.021 (0.015)	$\sigma_{3,2}$	0.75	0.042 (0.035)
	$\sigma_{4,1}$	1	0.018 (0.014)	$\sigma_{4,2}$	0.75	0.041 (0.029)
	$\sigma_{5,1}$	1	0.018 (0.017)	$\sigma_{5,2}$	0.75	0.042 (0.029)
	$\sigma_{6,1}$	1	0.028 (0.020)	$\sigma_{6,2}$	0.75	0.048 (0.035)
	$\sigma_{7,1}$	1	0.021 (0.015)	$\sigma_{7,2}$	0.75	0.045 (0.031)
	$\sigma_{8,1}$	1	0.022 (0.016)	$\sigma_{8,2}$	0.75	0.041 (0.031)
	$\sigma_{9,1}$	1	0.019 (0.016)	$\sigma_{9,2}$	0.75	0.047 (0.038)
	$\sigma_{10,1}$	1	0.021 (0.016)	$\sigma_{10,2}$	0.75	0.045 (0.034)
	$\beta_{1,1,1}$	-1	0.084 (0.032)	$\beta_{2,1,1}$	1	0.044 (0.034)
	$\beta_{1,1,2}$	1	0.025 (0.017)	$\beta_{2,1,2}$	-1	0.061 (0.040)
	$\beta_{1,2,1}$	1	0.046 (0.027)	$\beta_{2,2,1}$	-1	0.039 (0.031)
	$\beta_{1,2,2}$	0	0.030 (0.024)	$\beta_{2,2,2}$	-2	0.056 (0.039)
	$\Sigma_{\omega_{1,1,1}}$	0.25	0.018 (0.012)	$\Sigma_{\omega_{1,1,2}}$	0.56	0.034 (0.025)
	$\Sigma_{\omega_{1,2,1}}$	0.04	0.014 (0.011)	$\Sigma_{\omega_{1,2,2}}$	0.25	0.023 (0.017)
	$\Sigma_{\omega_{2,2,1}}$	0.25	0.024 (0.017)	$\Sigma_{\omega_{2,2,2}}$	0.56	0.036 (0.024)
	$\Sigma_{\xi_{1,1,1}}$	1	0.123 (0.079)	$\Sigma_{\xi_{1,1,2}}$	1	0.088 (0.063)
	$\Sigma_{\xi_{1,2,1}}$	0.25	0.193 (0.090)	$\Sigma_{\xi_{1,2,2}}$	0.25	0.052 (0.042)
	$\Sigma_{\xi_{1,3,1}}$	0.25	0.085 (0.056)	$\Sigma_{\xi_{1,3,2}}$	0.25	0.055 (0.037)
	$\Sigma_{\xi_{1,4,1}}$	0.25	0.086 (0.063)	$\Sigma_{\xi_{1,4,2}}$	0.25	0.059 (0.041)
	$\Sigma_{\xi_{2,2,1}}$	1	0.214 (0.123)	$\Sigma_{\xi_{2,2,2}}$	1	0.078 (0.058)
	$\Sigma_{\xi_{2,3,1}}$	0.25	0.099 (0.068)	$\Sigma_{\xi_{2,3,2}}$	0.25	0.057 (0.042)
	$\Sigma_{\xi_{2,4,1}}$	0.25	0.088 (0.066)	$\Sigma_{\xi_{2,4,2}}$	0.25	0.054 (0.042)
	$\Sigma_{\xi_{3,3,1}}$	1	0.081 (0.058)	$\Sigma_{\xi_{3,3,2}}$	1	0.079 (0.061)
	$\Sigma_{\xi_{3,4,1}}$	0.25	0.069 (0.048)	$\Sigma_{\xi_{3,4,2}}$	0.25	0.057 (0.040)
	$\Sigma_{\xi_{4,4,1}}$	1	0.084 (0.057)	$\Sigma_{\xi_{4,4,2}}$	1	0.069 (0.057)

TABLE 5.3 – Frequencies of selected number of mixture components based on the proposed BIC, for each true number of components.

		Minimal BIC_{MLFA}				
		$C = 1$	$C = 2$	$C = 3$	$C = 4$	Total
Correctly specified true model	$C = 1$	100	0	0	0	100
	C=2	10	75	13	2	100

intensities (low, moderate, high, severe). The pain mapping tool used to assess pain surface can be seen in figure 5.2. These 2 indicators were included as the outcomes matrix Y. From the 31 analysed items, 5 were removed as they were evaluating redundant outcomes (e.g. pain intensity is evaluated in an item of the ODI, the EQ-5D and the NPRS). In this analysis, we used data from 192 patients with at least 1 visit and no missing values in the outcomes described above.
The optimal number of classes was identified based on BIC_{MLFA} criterion. In the structural model, the latent factors were modeled using a quadratic function of time as fixed effects. Random slopes were considered for all the variables (quadratic time function) included in the fixed effect covariates. In the measurement model, we supposed a 4-factor structure for the loading matrices. Based on our previous work [28], we assumed that the following constructs represent the shared variance of the measured outcomes: pain intensity, psychological distress, disability and extent of pain. Therefore, we fixed one of the loadings associated with each construct to 1.

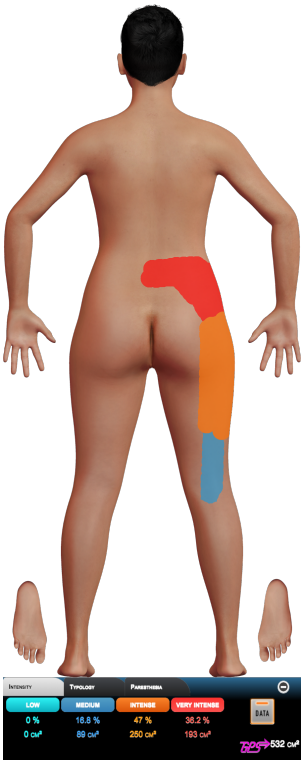

FIGURE 5.2 – An exemple of a pain surface evaluated using the pain mapping software.

The relationship between the posterior class membership and the biopsychosocial patient characteristics was tested using a χ^2 independance test for categorical characteristics or a t-test/Mann-Whitney test (depending on the normality of variables tested using a Shapiro-Wilk test) for continuous variables. Patients characteristics included age, sex, educational level, coping strategies, work status and pain duration.
The factor loadings of the proposed model are presented in Table 5.5. The fixed and random effects results are presented in Table 5.4. The factor loadings of the four latent factors extracted from this factor analysis can be found in Table 5.4. A two-components model was identified using the BIC_{MVCRE} criterion ($BIC_{MVCRE} = -10673.71$ for $C = 1$, $BIC_{MVCRE} = -12124.19$ for $C = 2$, $BIC_{MVCRE} = -10571.29$ for $C = 3$ and $BIC_{MVCRE} = -8555.66$ for $C = 4$). The percentage of patients in each component is 29.7% ($n = 57$) and 70.3% ($n = 135$) for the first and the second clusters respectively. We can observe that, for the first component, the first latent factor η_{11} mainly explains the variables "EQ-5D mobility", "EQ-5D usual activities", "ODI social life", "HADS enjoying things you used to enjoy", "HADS laughing and seeing the good side of things", "HADS feeling cheerful", "HADS feeling slowed down" and "HADS looking forward with enjoyment to things". These variables represent the subjective perception of the enjoyment that social and daily activities confer to the patient. On the other hand, the first factor η_{12} of the second majority cluster explains the items "Mobility", "Self-care" and "Usual activities" from the EQ-5D questionnaire, the items "Lifting", "Walking", "Standing", "social life" and "travelling" from the ODI questionnaire and the items "Enjoying thing you used to enjoy" from the HADS questionnaire. The majority of items in this factor are part of the EQ-5D and ODI questionaires which are used to evaluate functional disability. The second factor η_{21} regroups the items from the HADS questionnaire (anxiety and depression items) and the "Social life" and "Sex life" items from the ODI questionnaire. These two factors (η_{21} and η_{22}), which might represent patients psychological state, had a similar structure between the two mixture components. For patients in the first cluster, The third factor η_{31} comprises the items: "NPRS" (pain intensity), pain surface, four items from the HADS questionnaire and item "Sleeping" from the ODI questionnaire. However, the third factor η_{32} of the second cluster includes the items "pain surface", "ODI sitting" and the items "Self care", "Usual activities" and "personal care" from the EQ-5D questionnaire but no item from the HADS questionaire contrary to the first cluster. This reinforces that patients in the second class have a more objective body perception since their pain surface is associated with more objective outcomes ("functional capacity") than patients in the first cluster. The forth factors η_{41} and η_{42} include pain intensity ("NPRS") and items from the HADS questionnaire which suggests that pain intensity perception is associated with the psychological/emotional components for patients in both clusters. The correlations between the random effects can be found in Table 5.5. For the first cluster data, the random intercepts of the latent factors η_{11} and η_{21} are negatively correlated ($\rho(\xi_{0,1,1}, \xi_{0,2,1}) = -0.731$). Similarly, the random intercepts of the factors η_{31} and η_{41} are also negatively correlated ($\rho(\xi_{0,3,1}, \xi_{0,4,1}) = -0.873$). These correlations suggest that these latent outcomes might also come from a larger construct which might represent the global suffering. In addition, it is known that the psychological component of pain is associated with pain perception and the subjective perception of ones capability to accomplish daily life tasks, which might explain these correlations between these factors. On the other hand,

TABLE 5.4 – This table shows the factor loadings matrix of the four latent factors for each latent component. Each line corresponds to an observer/measured variable. Values in bold denote the observed variables which are loaded in the factor (higher coefficients). Positive values indicate that the factor is positively correlated with the observed variable while negative values indicate that the factor is negatively correlated with loading.

		Component 1 factors ($n = 57$)				Component 2 factors ($n = 135$)			
Variable name	Parameter	η_{11}	η_{21}	η_{31}	η_{41}	η_{12}	η_{22}	η_{32}	η_{42}
EQ5D Mobility	$\lambda_{1,.}$	**1**	0.09	0.38	0.44	1	0.33	0.50	0.05
EQ5D Self care	$\lambda_{2,.}$	0.54	-0.18	0.52	0.19	**0.78**	0.46	**0.77**	0.19
EQ5D Usual activities	$\lambda_{3,.}$	**0.90**	0.45	0.41	0.66	**0.81**	0.45	**0.75**	0.18
ODI Personal care	$\lambda_{4,.}$	0.58	0.13	0.54	0.41	**0.76**	0.49	**0.71**	0.21
ODI Lifting	$\lambda_{5,.}$	0.42	0.27	0.32	0.36	**0.78**	0.35	0.59	-0.15
ODI Walking	$\lambda_{6,.}$	0.42	0.06	0.57	0.54	**1.12**	0.25	0.02	-0.29
ODI Sitting	$\lambda_{7,.}$	0.23	0.36	0.19	0.50	0.27	0.08	**1.76**	-0.00
ODI Standing	$\lambda_{8,.}$	0.47	0.40	0.37	0.52	**0.84**	0.26	0.60	-0.05
ODI Sleeping	$\lambda_{9,.}$	0.23	0.17	**0.85**	0.52	0.25	0.17	**0.73**	0.26
ODI Sex life	$\lambda_{10,.}$	0.68	**0.75**	0.43	0.62	0.58	0.61	0.54	0.36
ODI Social life	$\lambda_{11,.}$	**0.85**	**0.94**	0.46	**0.73**	**0.78**	**0.84**	0.64	0.37
ODI Travelling	$\lambda_{12,.}$	0.67	0.51	0.42	0.49	**0.80**	0.44	**0.86**	-0.08
HAD Feeling tense or wound up	$\lambda_{13,.}$	0.21	1	0.52	**0.96**	0.11	1	0.68	**1.21**
HAD Enjoying thing you used to enjoy	$\lambda_{14,.}$	**0.97**	**1.07**	**0.79**	**0.74**	**0.71**	**0.95**	0.26	**0.75**
HAD Feeling that something awful is about to happen	$\lambda_{15,.}$	0.52	**0.96**	0.22	**0.85**	0.45	**1.22**	0.07	**1.09**
HAD Laughing and seeing the good side of thing	$\lambda_{16,.}$	**0.99**	**1.43**	**0.71**	**0.92**	0.49	**1.13**	0.34	**0.98**
HAD Worrying thoughts going through your mind	$\lambda_{17,.}$	0.48	**1.05**	0.45	**1.10**	0.28	**1.07**	0.29	**1.34**
HAD Feeling cheerful	$\lambda_{18,.}$	**0.81**	**1.16**	0.48	0.59	0.38	**1.11**	0.44	**1.10**
HAD Feeling slowed down	$\lambda_{19,.}$	**0.70**	**0.99**	**1.06**	**0.98**	0.50	**0.72**	0.36	**0.78**
HAD Frightened feeling like butterflies in the stomach	$\lambda_{20,.}$	0.45	**1.07**	0.12	**1.04**	0.37	**1.21**	0.21	**1.17**
HAD Looking forward with enjoyment to things	$\lambda_{21,.}$	**0.90**	**1.20**	**0.71**	0.64	0.51	**1.07**	0.37	**0.80**
HAD Sudden feeling of panic	$\lambda_{22,.}$	0.432	**0.86**	-0.18	**0.86**	0.29	**1.20**	0.13	**1.21**
HAD Enjoying a good book or radio/TV program	$\lambda_{23,.}$	0.38	**0.94**	**0.82**	0.34	0.22	0.67	0.23	**0.72**
Pain surface	$\lambda_{24,.}$	-0.06	0.00	**1**	0.54	0.13	0.02	**1**	0.68
Numeric pain rating scale	$\lambda_{25,.}$	0.46	0.58	**1.02**	**1**	0.55	0.39	0.63	**1**

TABLE 5.5 – This table shows the estimated parameters of the structural model (multivariate mixed effect model) in the real data example. As a reminder, this model estimates the evolution of the four latent factors as a quadratic function of time for each class. The table contains the probability of class membership for each class (π_1 and π_2), the fixed effects estimates (β coefficients where β_{c0f} represents the intercept for factor f in class c and β_{ct^pf} represents the regression coefficient for factor f in class c for the time degree p) and the correlation between the random effects. $\rho(\xi_{p,f,c}, \xi_{q,j,c})$ represents the correlation between the random effect p in the f^{th} latent factor and the random effect q in the j^{th} latent factor.

Component 1 parameters		Component 2 parameters	
Parameter	Estimation	Parameter	Estimation
π_1	0.296	π_2	0.703
$\beta_{1,0,1}$	0.422	$\beta_{2,0,1}$	0.223
$\beta_{1,0,2}$	-1.156	$\beta_{2,0,2}$	-1.229
$\beta_{1,0,3}$	0.014	$\beta_{2,0,3}$	-0.518
$\beta_{1,0,4}$	-0.573	$\beta_{2,0,4}$	-0.709
$\beta_{1,t,1}$	0.493	$\beta_{2,t,1}$	0.740
$\beta_{1,t,2}$	0.215	$\beta_{2,t,2}$	0.154
$\beta_{1,t,3}$	0.103	$\beta_{2,t,3}$	0.453
$\beta_{1,t,4}$	-0.559	$\beta_{2,t,4}$	-0.872
$\beta_{1,t^2,1}$	1.033	$\beta_{2,t^2,1}$	0.943
$\beta_{1,t^2,2}$	-0.739	$\beta_{2,t^2,2}$	-1.149
$\beta_{1,t^2,3}$	-0.029	$\beta_{2,t^2,3}$	0.018
$\beta_{1,t^2,4}$	0.828	$\beta_{2,t^2,4}$	1.029
$\rho(\xi_{0,1,1}, \xi_{1,1,1})$	0.556	$\rho(\xi_{0,1,2}, \xi_{1,1,2})$	-0.369
$\rho(\xi_{0,1,1}, \xi_{2,1,1})$	0.581	$\rho(\xi_{0,1,2}, \xi_{2,1,2})$	-0.375
$\rho(\xi_{1,1,1}, \xi_{2,1,1})$	0.976	$\rho(\xi_{1,1,2}, \xi_{2,1,2})$	0.986
$\rho(\xi_{0,2,1}, \xi_{1,2,1})$	-0.802	$\rho(\xi_{0,2,2}, \xi_{1,2,2})$	-0.467
$\rho(\xi_{0,2,1}, \xi_{2,2,1})$	0.883	$\rho(\xi_{0,2,2}, \xi_{2,2,2})$	0.881
$\rho(\xi_{1,2,1}, \xi_{2,2,1})$	-0.777	$\rho(\xi_{1,2,2}, \xi_{2,2,2})$	-0.555
$\rho(\xi_{0,3,1}, \xi_{1,3,1})$	0.174	$\rho(\xi_{0,3,2}, \xi_{1,3,2})$	-0.791
$\rho(\xi_{0,3,1}, \xi_{2,3,1})$	-0.332	$\rho(\xi_{0,3,2}, \xi_{2,3,2})$	-0.754
$\rho(\xi_{1,3,1}, \xi_{2,3,1})$	-0.912	$\rho(\xi_{1,3,2}, \xi_{2,3,2})$	0.943
$\rho(\xi_{0,4,1}, \xi_{1,4,1})$	0.358	$\rho(\xi_{0,4,2}, \xi_{1,4,2})$	0.965
$\rho(\xi_{0,4,1}, \xi_{2,4,1})$	-0.400	$\rho(\xi_{0,4,2}, \xi_{2,4,2})$	-0.971
$\rho(\xi_{1,4,1}, \xi_{2,4,1})$	-0.977	$\rho(\xi_{1,4,2}, \xi_{2,4,2})$	-0.987
$\rho(\xi_{0,1,1}, \xi_{0,2,1})$	-0.731	$\rho(\xi_{0,1,2}, \xi_{0,2,2})$	0.072
$\rho(\xi_{0,1,1}, \xi_{0,3,1})$	0.146	$\rho(\xi_{0,1,2}, \xi_{0,3,2})$	0.268
$\rho(\xi_{0,1,1}, \xi_{0,4,1})$	-0.233	$\rho(\xi_{0,1,2}, \xi_{0,4,2})$	0.351
$\rho(\xi_{0,2,1}, \xi_{0,3,1})$	-0.081	$\rho(\xi_{0,2,2}, \xi_{0,3,2})$	0.696
$\rho(\xi_{0,2,1}, \xi_{0,4,1})$	0.122	$\rho(\xi_{0,2,2}, \xi_{0,4,2})$	0.845
$\rho(\xi_{0,3,1}, \xi_{0,4,1})$	-0.873	$\rho(\xi_{0,3,2}, \xi_{0,4,2})$	0.701
$\rho(\xi_{1,1,1}, \xi_{1,2,1})$	0.647	$\rho(\xi_{1,1,2}, \xi_{1,2,2})$	0.451
$\rho(\xi_{1,1,1}, \xi_{1,3,1})$	0.778	$\rho(\xi_{1,1,2}, \xi_{1,3,2})$	0.988
$\rho(\xi_{1,1,1}, \xi_{1,4,1})$	-0.949	$\rho(\xi_{1,1,2}, \xi_{1,4,2})$	-0.953
$\rho(\xi_{1,2,1}, \xi_{1,3,1})$	0.844	$\rho(\xi_{1,2,2}, \xi_{1,3,2})$	0.521
$\rho(\xi_{1,2,1}, \xi_{1,4,1})$	-0.820	$\rho(\xi_{1,2,2}, \xi_{1,4,2})$	-0.658
$\rho(\xi_{1,3,1}, \xi_{1,4,1})$	-0.864	$\rho(\xi_{1,3,2}, \xi_{1,4,2})$	-0.972
$\rho(\xi_{2,1,1}, \xi_{2,2,1})$	-0.999	$\rho(\xi_{2,1,2}, \xi_{2,2,2})$	-0.999
$\rho(\xi_{2,1,1}, \xi_{2,3,1})$	-0.984	$\rho(\xi_{2,1,2}, \xi_{2,3,2})$	0.975
$\rho(\xi_{2,1,1}, \xi_{2,4,1})$	0.999	$\rho(\xi_{2,1,2}, \xi_{2,4,2})$	0.999
$\rho(\xi_{2,2,1}, \xi_{2,3,1})$	0.985	$\rho(\xi_{2,2,2}, \xi_{2,3,2})$	-0.975
$\rho(\xi_{2,2,1}, \xi_{2,4,1})$	-0.999	$\rho(\xi_{2,2,2}, \xi_{2,4,2})$	-0.999
$\rho(\xi_{2,3,1}, \xi_{2,4,1})$	-0.985	$\rho(\xi_{2,3,2}, \xi_{2,4,2})$	0.975

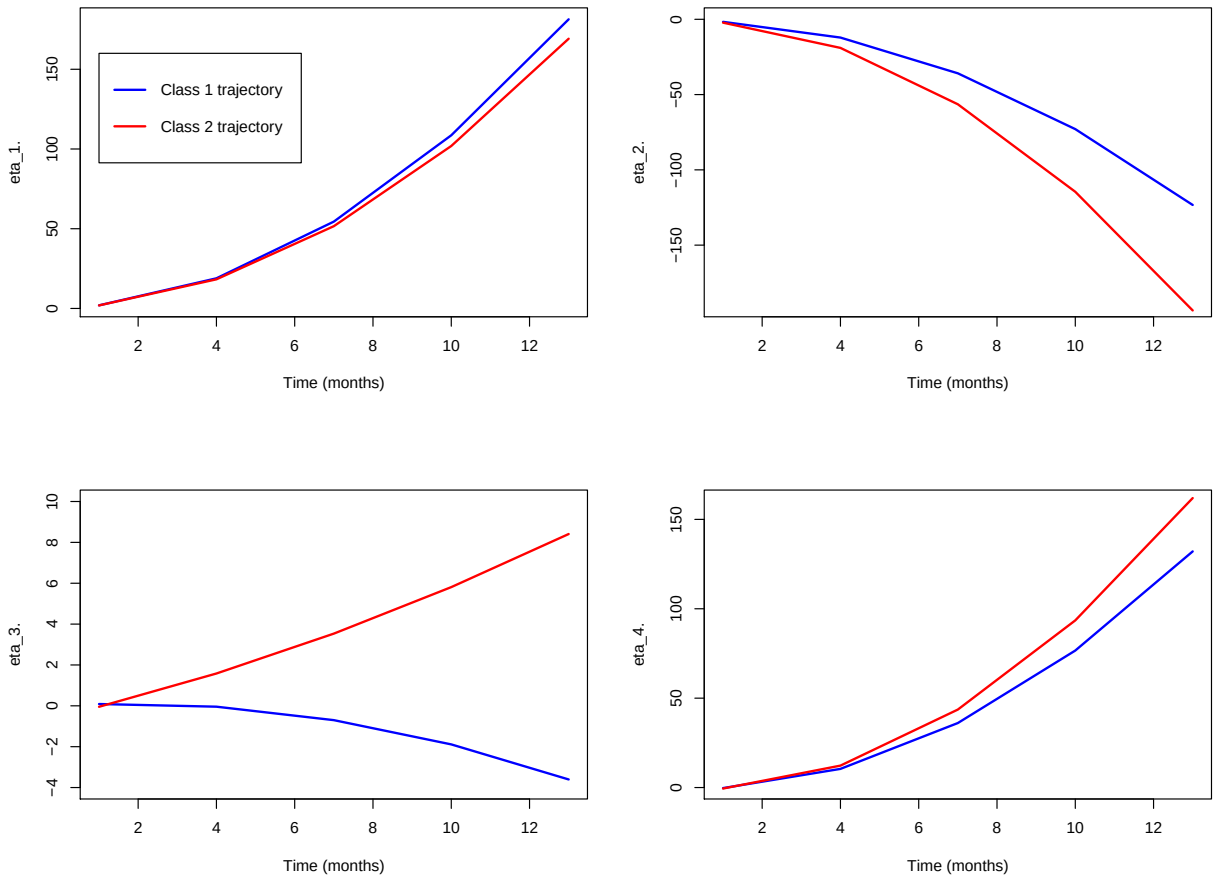

FIGURE 5.3 – Time trajectories of the four extracted latent factors for class 1 (blue lines) and class 2 (red lines) patients.

for cluster 2 patients, high positive correlations were observed between the random intercepts of the fourth factor η_{42} and both the second and third factors (η_{22} and η_{32}). The time trajectories of the latent factors of class 1 and class 2 patients can be found in Figure 5.3.

5.6.1 Variables associated with class membership

In this section, we studied the link between age, sex, educational level, coping strategies, work status, pain duration and baseline ODI, NPRS and HADS and the identified clusters. Bivariate analysis results are presented in table 5.6. The bivariate analysis revealed that patients in the first cluster had a lower NPRS than patients in the second cluster (5.6 (1.4) vs 6.3 (1.5), $p = 0.002$). ODI score at baseline also had a statistically significant impact on class membership (31.2 (10.0) vs 35.4 (12.3), $p = 0.0002$). The mean diversion CSQ scores was higher in the second cluster than in the first cluster but it was only significant at a 10% level (11.6 (3.5) vs 12.6 (3.7), $p = 0.08$). Age was also statistically different between the two clusters at a 10% significance level (50.2 (13.2) vs 53.6 (12.3), $p = 0.08$). Sex, HADS score,

Variable	Mean (sd)/n(%)		
	Class 1 $n = 57$	Class 2 $n = 135$	p-value
global NPRS at baseline	5.6 (1.4)	6.3 (1.5)	0.002
ODI percentage at baseline	31.2 (10.0)	35.4 (12.3)	0.0002
HADS total score at baseline	18.2 (7.7)	18.7 (6.5)	0.59
Sex (male)	24/57 (42.1%)	60/135 (44.4%)	0.87
Age (years)	50.2 (13.2)	53.6 (12.3)	0.08
Current pain duration (years)	3.5 (4.4)	4.9 (6.6)	0.12
Work social category			0.34
Executive	1 (1.8%)	8 (5.9%)	
White-collar	32 (56.1%)	56 (41.5%)	
Blue-collar	14 (24.6%)	47 (34.8%)	
Intermidiate profession	6 (10.5%)	13 (9.6%)	
Other	4 (7.0%)	11 (8.1%)	
Level of study (years)	11.1 (2.7)	10.7 (4.0)	0.59
CSQ catastrophizing score*	13.2 (4.0)	14.1 (4.8)	0.36
CSQ divertion score*	11.6 (3.5)	12.6 (3.7)	0.08
CSQ ignorance score*	8.7 (3.1)	9.6 (3.5)	0.14

TABLE 5.6 – Demographic and cognitivo-behavioral factors associated with class membership. *CSQ catastrophizing: the score of the Coping Strategies Questionnaire in the catastrophizing dimension (i.e. exagerating pain sensation and intensity). CSQ diversion: the score of the diversion dimension of the CSQ (i.e. diverting attention from pain using external stimulus, as a strategy to cope with pain). CSQ ignorance: the Score of the ability of the patient to ignore pain in order to cope with it.

CSQ catastrophizing and ignorance scores, pain duration, work social category and level of study had no statistically significant impact on class membership ($p = 0.87$ for sex, $p = 0.59$ for HADS score, $p = 0.12$ for pain duration, $p = 0.34$ for work social category, $p = 0.59$ for level of study, $p = 0.36$ for CSQ catastrophizing score and $p = 0.14$ for CSQ ignorance score).

5.7 Discussion

In this chapter, we have proposed a mixture of factor analysis models which can be used to conduct a cluster-based, longitudinal factor analysis. The proposed model can be used to cluster individuals based on the associations between several different variables in an unsupervised manner. It can also explain the relationship between the cluster-specific longitudinal latent factors and a set of explanatory variables using a mixed effects model. The model was estimated using the EM algorithm. We have also proposed a BIC criterion which can be used to identify the optimal number of mixture components. Our simulations showed that the EM algorithm is able to estimate the true parameters under some identifiability assumptions, for both balanced and unbalanced data. The advantages of

using a longitudinal factor analysis mixed model instead of a SEM or a two-step approach are discussed in the paper by An et al. [7]. These advantages include the ability to model data where subjects are measured at different time points and accepting covariates other than time for the random effects.
Regarding the mixture of factor analyzer approaches, several authors have proposed similar methods [163–165]. However, to our knowledge, the method proposed in section **??** is the only one simultaneously taking into account both intra-subject correlation and inter-subject unobserved heterogeneity, which makes it more adequate for the factor analysis of multi-level or longitudinal heterogeneous data. The proposed model also has the advantage that the mixture/heterogeneity can be incorporated into all the parameters of the model including the uniqueness (residual variances) and the variance/covariance components of the mixed effects model. Occasionally, only the heterogeneity in the factor structure (loading matrices) is of interest. This can be evaluated using our model since some parameters can be easily fixed between clusters. In addition, from a practical point of view, the existing methods for clustering observations based on a factor model can only consider observations independently and therefore, two different observations from the same subject may be attributed to two different clusters. Another point which was discussed in this chapter is the possibility of evaluating measurement invariance based on the optimal number of classes. Measurement invariance can be investigated by comparing the 1-class model (i.e. invariance assumption) to the models with more than two classes using goodness of fit measures such as the BIC criterion proposed in section **??**.
In this chapter, we assumed that the number of latent factors is similar so as to simplify the estimation algorithm implementation and interpretability of the model. However, the number of factors can easily be made to vary between classes. The factor loading matrix is estimated "independently" for each class so it is only necessary to specify the structure of the factor loadings for each model with a different number of columns for the loading matrices. It can also be noted that in the equations used to estimate the model, the parameters of each class are estimated using only the remaining parameters of that same class. Only the mixture parameter π requires knowing the parameters of all the classes but the estimation of π requires only the likelihoods of the parameters of each class, which are also calculated independently. However, taking a different number of factors for each class changes the equation of the proposed BIC_{MLFA} because it is no longer sufficient to multiply the number of classes by the number of model parameters within the mixture. When the number of latent factors changes, the number of parameters for each class must be calculated separately, and then the sum of the number of parameters of all the classes must be calculated, which is likewise easy to do.
The proposed model also offers another degree of flexibility since we can use different observed variables in each class. For example, let's say we have j variables and we hypothesize that our sample contains two latent sub-samples which need to be modeled by different variables, then we can make a mixture of factorial analyses including variables Y_1, Y_2, ..., Y_k for the first class and other variables Y_{k+1}, ..., Y_j for the second class. However, we think that it must be more complicated to interpret the factors when this is done in an exploratory manner. In practice, it is better to have a similar number of factors because it facilitates the comparison between the latent classes. On the other hand, it can be interesting to vary the number of factors when we suspect that one of the groups needs a

different number of factors to better describe their factor structure.
Despite the advantages of the proposed model, several limitations and extensions need to be discussed. The mixture of factor analyzers proposed here accepts only continuous outcomes, which limits its applicability. Classical factor analysis is generally applied to questionnaires consisting of several Likert scales which are ordinal. An extension including methods such as latent responses [166, 167] (where the categorical variable is assumed to be a dichotomization using unknown cut-offs of a latent continuous variable) can allow categorical outcomes to be considered in the model. This method replaces the categorical outcomes by a latent continuous variable which is estimated before being included in the factor analysis. This adds a new level to the proposed model which would render it computationally intensive since the E-step in the EM algorithm no longer has an explicit expression and only an approximation using numerical methods is possible. Another limitation of the EM algorithm is that it can converge to local maxima which might lead to erroneous interpretations, specially in the case of such complex models. To address this issue, we used the multistart method, which corresponds to estimating the model using several random initial values and finally conserving the model which yields the highest likelihood (i.e. closest to the global maxima). In addition, the EM algorithm has a slow convergence rate. Several techniques to accelerate the convergence of the EM algorithm have been proposed in the literature. These methods include the Expectation Conditional-Maximization Either (ECME) [168], the Minorization-Maximization (MM) [169] and the Parameter Expansion EM (PX-EM) algorithms [170]. One final extension worth discussing is the inclusion of a temporal component in the estimation of the factor model. A model where the factor loadings could also vary with time could be of interest to study the temporal invariance of the factorial structure. Several methods have been proposed to account for the temporal changes in the loading matrix in the econometric research and psychological process studies due to their high temporal variability. These methods include modeling the current factor loadings using an autoregressive process of the past factor loadings (dynamic factor models) [171]. Other methods include the use of non-parametric estimation to model the factor loadings as a smooth function of time [172]. In relation to our real data example, a time-varying factor analysis model could help identify the structural changes over time in the relationship between the different pain components and therefore, evaluate the relevance of the biopsychosocial model of pain and the fear-avoidance model [173], which were developed in the 70's, but their evolution over time has never been considered.
We think that the proposed model could help to a great extent in the evaluation of chronic pain patients. As discussed before, current pain evaluation based on pain intensity alone leads to disastrous consequences [17]. These consequences could be avoided if pain is evaluated more adequately using more complex tools which consider both the patient-specific characteristics and the multidimensional components of a person's pain.

5.A Appendix for this chapter

For the case where no mixture is included in the model, the conditional expectations are detailed in the paper by An et al. [7] from which we derive the expectations in the case of a mixture of longitudinal factor analysis models. To calculate the conditional expectation for

the sufficient statistics in algorithm 3, we get the joint distribution of $(y_i, \eta_{ic}, \xi_{ic})$ conditional on v_{ic}.

$$\begin{pmatrix} y_i \\ \eta_{ic} \\ \xi_{ic} \end{pmatrix} \Bigg| v_{ic} \sim N\left(\begin{pmatrix} I_{t_i} \otimes \Lambda_c x_i \beta_c \\ x_i \beta_c \\ 0 \end{pmatrix}, \begin{pmatrix} \sum_{y_i} & \sum_{y_i \eta_{ic}} & \sum_{y_i \xi_{ic}} \\ \sum_{\eta_{ic} y_i} & \sum_{\eta_{ic}} & \sum_{\eta_{ic} \xi_{ic}} \\ \sum_{\xi_{ic} y_i} & \sum_{\xi_{ic} \eta_{ic}} & \sum_{\xi_{ic}} \end{pmatrix} \right), \tag{5.14}$$

where

$$\Sigma_{\eta_{ic}|v_{ic}} = z_i \Sigma_{\xi_c} (z_i)^T + I_{t_i} \otimes \Sigma_{\omega_c}, \tag{5.15}$$

$$\Sigma_{y_i|v_{ic}} = (I_{t_i} \otimes \Lambda_c) \Sigma_{\eta_{ic}|v_{ic}} (I_{t_i} \otimes \Lambda_c)^T + I_{t_i} \otimes diag(\sigma_{1c}^2, \ldots, \sigma_{Jc}^2), \tag{5.16}$$

$$\Sigma_{\xi_{ic}|v_{ic}} = \Sigma_{\xi_c}, \tag{5.17}$$

$$\Sigma_{y_i \eta_{ic}|v_{ic}} = (I_{t_i} \otimes \Lambda_c) \Sigma_{\eta_{ic}|v_{ic}}, \tag{5.18}$$

$$\Sigma_{y_i \xi_{ic}|v_{ic}} = (I_{t_i} \otimes \Lambda_c)(z_i \Sigma_{\xi_c}), \tag{5.19}$$

$$\Sigma_{\eta_{ic} \xi_{ic}|v_{ic}} = z_i \Sigma_{\xi_c}. \tag{5.20}$$

Therefore, ξ_{ic} and η_{ic} conditional on y_i and v_{ic} are normally distributed with the following mean and covariance components.

$$\mu_{\eta_{ic}|y_i, v_{ic}} = x_i \beta_c + \Sigma_{\eta_i y_i|v_{ic}} \Sigma_{y_i|v_{ic}}^{-1} (y_i - I_{t_i} \otimes \Lambda_c x_i \beta_c), \tag{5.21}$$

$$\Sigma_{\eta_{ic}|y_i, v_{ic}} = \Sigma_{\eta_{ic}|v_{ic}} - \Sigma_{\eta_{ic} y_i|v_{ic}} \Sigma_{y_i|v_{ic}}^{-1} \Sigma_{y_i \eta_{ic}|v_{ic}}, \tag{5.22}$$

$$\mu_{\xi_{ic}|y_i, v_{ic}} = \Sigma_{\xi_{ic} y_i|v_{ic}} \Sigma_{y_i|v_{ic}}^{-1} (y_i - I_{t_i} \otimes \Lambda_c x_i \beta_c), \tag{5.23}$$

$$\Sigma_{\xi_{ic}|y_i, v_{ic}} = \Sigma_{\xi_{ic}|v_{ic}} - \Sigma_{\xi_{ic} y_i|v_{ic}} \Sigma_{y_i|v_{ic}}^{-1} \Sigma_{y_i \xi_{ic}|v_{ic}}, \tag{5.24}$$

The joint variable (η_{ic}, ξ_{ic}) conditional on y_i and v_{ic} is also normally distributed with the covariance matrix:

$$\Sigma_{\eta_{ic}, \xi_{ic}|y_i, v_{ic}} = \begin{pmatrix} \Sigma_{\eta_{ic}|y_i, v_{ic}} & \Sigma_{\eta_{ic} \xi_{ic}|y_i, v_{ic}} \\ \Sigma_{\xi_{ic} \eta_{ic}|y_i, v_{ic}} & \Sigma_{\xi_{ic}|y_i, v_{ic}} \end{pmatrix} = \begin{pmatrix} \Sigma_{\eta_{ic}|v_{ic}} & \Sigma_{\eta_{ic} \xi_{ic}|v_{ic}} \\ \Sigma_{\xi_{ic} \eta_{ic}|v_{ic}} & \Sigma_{\xi_{ic}|v_{ic}} \end{pmatrix} - \begin{pmatrix} \Sigma_{\eta_{ic} y_i|v_{ic}} \\ \Sigma_{\xi_{ic} y_i|v_{ic}} \end{pmatrix} \Sigma_{y_i|v_{ic}}^{-1} (\Sigma_{y_i \eta_{ic}|v_{ic}}, \Sigma_{y_i \xi_{ic}|v_{ic}}), \tag{5.25}$$

The conditional covariance of $\eta_{ic}\xi_{ic}$ on y_i and v_{ic} is:

$$\Sigma_{\eta_{ic}\xi_{ic}|y_i,v_{ic}} = \Sigma_{\eta_{ic}\xi_{ic}} - \Sigma_{\eta_{ic}y_i|v_{ic}}\Sigma^{-1}_{y_i|v_{ic}}\Sigma_{y_i\xi_{ic}|v_{ic}}, \tag{5.26}$$

We start by calculating:

$$\mathbb{E}(v_{ic}|y_i) = \frac{\pi_c\phi_{c,obs}(y_i, E_c, V_c)}{\sum_{l=1}^{C}\pi_c\phi_{c,obs}(y_i, E_c, V_c)} \tag{5.27}$$

where

$$\phi_{c,obs}(y_i, E_{ic}, V_{ic}) = \prod_{t=1}^{t_{n_i}}(\frac{1}{\sqrt{\det(V_{ic})}}\exp(-\frac{1}{2}(y_{it} - E_{ic})^T V^{-1}(y_{it} - E_{ic}))$$

and $E_{ic} = \Lambda_c X_{itc}\beta_c$ and $V_{ic} = \Lambda_c Z_{itc}\Sigma_{\xi_c}Z^T_{itc} + \Sigma_{\omega_c}\Lambda^T_c + diag(\sigma^2_{1c}, \ldots, \sigma^2_{Jc})$.

We then calculate the other sufficient statistics:

$$\mathbb{E}(\eta_{itc}|y_i, v_{ic}) = \mu_{\eta_{ic}|y_i,v_{ic}}[1 + (t-1) * d\colon t * d], \tag{5.28}$$

where the notation $[i\colon j]$, means the vector values from index i to index j. The notation $[i\colon j, k\colon l]$ means the submatrix including the lines from i to j and the columns from k to l.

$$\mathbb{E}(\eta_{itc}\eta^T_{itc}|y_i, v_{ic}) = \mathbb{E}(\eta_{itc}|y_i, v_{ic})\,\mathbb{E}(\eta_{itc}|y_i, v_{ic})^T + \Sigma_{\eta_{ic}|y_i,v_{ic}}[1+(t-1)*d\colon t*d,\ 1+(t-1)*d\colon t*d], \tag{5.29}$$

$$\mathbb{E}(\xi_{ic}|y_i, v_{ic}) = \mu_{\xi_{ic}|y_i,v_{ic}}, \tag{5.30}$$

$$\mathbb{E}(\xi_{ic}\xi^T_{ic}|y_i, v_{ic}) = \mathbb{E}(\xi_{ic}|y_i, v_{ic})\,\mathbb{E}(\xi_{ic}|y_i, v_{ic})^T + \Sigma_{\xi_{ic}|y_i,v_{ic}}, \tag{5.31}$$

$$\mathbb{E}(\eta_{itc}\xi^T_{ic}|y_i, v_{ic}) = \mathbb{E}(\eta_{itc}|y_i, v_{ic})\,\mathbb{E}(\xi_{ic}|y_i, v_{ic})^T + \Sigma_{\eta_{ic}\xi_{ic}|y_i,v_{ic}}[1 + (t-1) * d\colon t * d,\ :]. \tag{5.32}$$

Chapitre 6

Conclusion and perspectives

6.1 Discussion and conclusion

In this thesis, we were able to propose a set of methods to analyze and model the relationship between several latent or observed variables and one or several outcomes. We developed these methods in an attempt to address several limitations of the current statistical methods used to analyze heterogeneous, correlated longitudinal data. Unfortunately, currently available methods in the literature for analyzing longitudinal data do not take optimally into account data heterogeneity (*i.e.* the study sample might be drawn from different latent subpopulations) nor the intra-subject correlation. We proposed in this manuscript several methods which although still have some limitations, which will be discussed later on, can be considered a step forward towards better alternatives for analyzing longitudinal data. The methods proposed in this manuscript were developed by merging several classical frameworks together to allow more flexible and relevant modeling of this type of data. The first model we proposed was achieved by merging the mixture of models framework with the varying coefficient models and the mixed effect models. This allowed us to develop the mixture of varying coefficient models with random effect processes. This model could be used to estimate the time-varying effects of one or several covariates on an observed outcome while taking into account both intra-subject correlation and population heterogeneity. Intra-subject correlation was considered using within-subject random effect processes where each subject effect is represented by an observation from a functional distribution. On the other hand the heterogeneity is considered by using the mixture of models framework where each latent cluster of patients is modeled by a different random varying coefficient model. The second model we proposed in this manuscript was the mixture of longitudinal factor analysis models. This was achieved by again merging the mixture of models framework with the factor analysis framework and the mixed effect models. This model was developed in order to study the factorial structure of multivariate latent outcomes in heterogeneous populations and to estimate the impact of observed covariates on these latent outcomes. We then proposed iterative algorithms to estimate these models' parameters. These models were then applied to both simulated data and real data to assess their applicability and their limits.

In order to get a first outlook on the applicability of these types of models and to illustrate the limits of the models proposed in the literature, we started by applying the mixture of mixed effect models [80] to data of 200 patients with chronic pain after spinal surgery [41]. This application allowed us to identify two classes of patients based on the impact of pain intensity, functional disability and psychological distress on health-related quality of life. The first "disability" class consisting of 68.7% of patients, represented patients with health-related quality of life which was predominantly impacted by functional disability and depression, while the second "pain intensity" class consisted of 31.3% of patients for whom the health-related quality of life was mainly impacted by pain intensity and depression. We also found that factors such as level of education, perceived arduous working conditions, anxiety/depression, and coping strategies were significantly associated with the class membership. Thanks to this work we showed one primary hypothesis of our work which is that the health-related quality of life of patients with chronic pain after spinal surgery is multidimensional and that the impact of different pain dimensions on patients' health-related quality of life is heterogeneous. In the literature, the multidimensionality and heterogeneity of chronic pain were assumed from a theoretical point of view by clinicians but were never demonstrated empirically.

The developments in this chapter helped us reconsider the evaluation of chronic pain patients. These reconsiderations included the validity of one of the most used health-related quality of life measures, which is the EQ-5D index (described in chapter 3). This questionnaire is used by health structures and researchers when evaluating the cost effectiveness of a new chronic pain therapy in order to consider its benefit for both the patient's health and the economy (*e.g.* reimbursement structures). The 5 dimensions of the EQ-5D includes pain intensity, mobility, self-care, usual activities, and anxiety/depression. Currently, a unique EQ-5D tool is used to evaluate the health-related quality of life in each country. Each of the 5 previously given dimensions is weighted similarly for chronic pain patients regardless of their sociodemographic and cognitive-behavioural characteristics. Our conclusion in chapter 3 puts into question the way health-related quality of life is evaluated. We concluded that the supposedly homogeneous population regarding the impact of pain and functional disability on health related quality of life is heterogeneous since two clusters of patients were identified (the first cluster for whom functional disability has more impact on quality of life and a second cluster for whom pain intensity is more impactful). We suggest that researchers need to develop a new health related quality of life measure that incorporates the intrinsic sociodemographic and cognitive-behavioural differences between patients. This would allow a better evaluation of the benefit of chronic pain therapies on the health-related quality of life. However, one of the factors that might influence the impact of pain intensity and functional disability on the quality of life of chronic pain patients is pain duration as supported by the medical literature [19–21, 174]. However, to be able to study the impact of time/pain duration on our clusters, a more flexible model needed to be developed. Accordingly, after we showed that chronic pain patients are heterogeneous in their way of responding to functional disability and pain intensity alterations, we wanted to evaluate if this heterogeneity varies over time as described in chapter 4.

For this purpose, we have introduced a new statistical model formulated as a mixture of varying-coefficient models with random effect processes. We have shown using simulated

data that this model gives estimates which are less biased and deviate less from the true estimates compared to the models proposed in the literature [85]. Our proposed model also allowed the estimation of time-varying smooth fixed effects and random effect processes thus becoming a "non-parametric" generalization of the mixture of mixed effect models applied in chapter 3. We also applied our model to real data allowing us to observe the changes over time in the effects of pain intensity, disability and psychological distress on health-related quality of life identified in chapter 3. We observed that the impact of pain intensity on health-related quality of life decreases over time for patients in both classes. This finding is consistent with the findings of Hashmi et al. [20] which found using functional Magnetic Resonance Imaging (MRI) data that patients with acute (short-term pain) had more brain activity in the circuitry associated with sensory acute pain while patients with chronic pain had more activity in the circuitry associated with affect/emotions. We also confirmed the heterogeneity of patients health-related quality of life. The conclusions obtained from the mixture of varying-coefficients models in chapter 4 were different from those obtained in chapter 3. The difference lies in the fact that using the MVCRE model, functional disability had important effects for both classes contrary to the mixture of mixed effects model in which the second class had non-significant effect of functional disability on health-related quality of life. However, the two classes in the MVCRE model had different impacts of pain intensity on health-related quality of life.

One of the limitations of our methodology in chapter 4 is that our study sample was exclusively drawn from the chronic pain after spinal surgery population (pain for more than 6 months), which cannot allow us to extend our findings to the general painful population including acute, sub-acute and chronic general (not necessary postoperative) pain. In order to address this limitation and to achieve other objectives, we have developed the ongoing study PREDIPAIN (funded by the "Fonds Aliénor" initiative created by the Poitiers University Hospital) in which patients with leg and back pain with chronic or subacute back and leg pain are included. This will help us to identify the differences between the subacute/chronic and virgin/postoperative back pain regarding the evolution and the impact of the sensory, emotional and functional components of pain on patients health.

A second limitation of the MVCRE model proposed in chapter 4 is that it can only be applied to continuous outcomes. An extension to generalized outcomes such as binary and survival outcomes is needed in order to broaden its scope of application. This can be achieved by replacing the classical varying-coefficient models for continuous outcomes by their generalized varying-coefficient models counterparts which have been proposed in the literature [57, 58]. These models do not have closed form solutions and therefore need to be estimated using numerical methods such as Newton-Raphson algorithm but this can be integrated easily in the proposed EM algorithm for estimating the MVCRE model. In addition, in order to estimate the random effect processes, a two-step procedure using a generalized mixed effect model followed by kernel local polynomial smoothing can be used. We are currently working on the generalization of the MVCRE model to other types of outcomes. In addition, we are also working on implementing the model in a more computationally efficient programming language. The implementation used in chapter 4 was written in `R` and `Python` which unfortunately are computationally slow. We are currently working on an implementation in `C++`, the final goal being to produce an R package including the time-varying mixture of varying-coefficient models with random

effect processes and their generalized counterpart.
Using data from 200 patients with post-operative chronic back pain, we showed in chapter 3 [41] and chapter 4 that the quality of life of different clusters of patients is impacted differently by pain intensity and by functional disability, suggesting that the structure of the relationship between different chronic pain dimensions vary among presumed homogeneous populations. We also found that the relationships between different pain dimensions increased with time. This also suggests that the structure relating different pain dimensions changes with time and between clusters of patients with specific characteristics.
In a paper by Rigoard et al (2021) [28], we proposed a composite evaluation index using data from the PREDIBACK study data. This composite index was developed using classical factor analysis on items from several questionnaires. A 2-factor model yielded the best results (based on the very simple structure criterion). These 2 factors representing functional disability and psychological distress were included with 2 other variables (pain intensity and pain surface) in a PCA. The first component in the PCA was taken as the new composite score. One of the limitations of this approach is that it considers chronic pain and its dimensions as both time and group invariant which is not the case in reality. For example, some patients in our studied population are in a wheelchair due to amputation, which means that functional capacity and specifically items evaluating dimensions such as walking and standing should not play an important role for these patients since no change would be observed even if the patient becomes pain free. To address these discussed limitations related to considering patients as similar when evaluating their pain, we developed a mixture of longitudinal factor analysis models with time-varying parameters. we have introduced a new extension of the factor analysis model that releases both time and group invariance restrictions. These restrictions are assumed in order to allow comparability between factor scores [175]. Measurement invariance (*i.e.* similar factor structures and loadings across groups), in the structural equation modeling framework, can be tested when the groups are known, using a Chi-squared difference test or goodness of fit indices comparing the metric invariance model with the unconstrained model [176, 177]. In our model the groups where the non-invariance is observed are latent and can not be compared directly. Therefore, the proposed model has the advantage of allowing the number of latent groups to vary using the mixture modeling framework where several methods have been proposed to identify the optimal number of groups [178, 179]. We have proposed a modified BIC criterion which could be used to identify the number of groups in our model. Group invariance could be assumed when the optimal number of classes is $C = 1$. Regarding time non-invariance in our model, it is taken into account using a polynomial function of time incorporated into the intercepts, factor loadings and factor scores. In our model, we might be able to evaluate the factors time non-invariance by evaluating the goodness of fit of our unrestricted model compared to the configural invariance model (when the overall factor structure is fixed), the metric invariance model (when the factor loadings are supposed to be equal among groups/in time) and the scalar invariance (when the intercepts are assumed to be equal in time and among groups).
This modeling framework seems adequate to model the structure relating the different chronic pain dimensions in order to achieve a precise patient-specific evaluation and to arrive at appropriate inferences on both how these structures differ between-groups and how they evolve over time.

6.2 Perspectives and future work

The mixture of longitudinal factor analysis model proposed and discussed above seems to be adequate for analyzing heterogeneous multivariate longitudinal data.
Although this model is flexible and has several strengths, its current form only accepts continuous outcomes which limits its applicability knowing that factor analysis are generally applied to questionnaires where items are represented by likert scales, which are ordinal. To be able to incorporate categorical variables in factor analysis, methods such as latent responses can be considered (the categorical variable is assumed to be a dichotomization of a latent continuous variable) [166, 167]. In the latent response framework, the latent continuous variable needs to be estimated before including it in the proposed factor analysis which adds a new level/equation in the proposed model. This new component of the model would render it computationally intensive since the E-step in the EM algorithm no longer has an explicit expression and only an approximation is possible (with the Monte-Carlo method for example). Similarly, the model proposed in chapter 4 also needs to be generalized to bivariate
As discussed before, the methods for analysing heterogeneous longitudinal data developed in this thesis will be included in an R package, which, potentially, would be deposited on `CRAN`. Currently, both the time-varying mixture of varying coefficient models with random effects and the mixture of longitudinal factor analysis models are developed in R which is computationally slow, specially when it comes to matrix calculations and iterative methods (loops). To overcome this limitation, we will rewrite all the algorithms and methods proposed in this thesis using the `C++` programming language. Furthmore, we would also like to include all the extensions discussed here in the final package. This extensions include proposing generalization of the time-varying mixtures of varying-coefficient models for categorical and time to event outcomes, developing statistical methods to test whether the time-varying coefficients are constant or not (*i.e.* whether we need to model the coefficient using smooth functions or if parametric constant coefficients are sufficient) and finally to extend the model to cases where multiple outcomes need to be modeled simultaneously. The mixture of longitudinal factor analysers could also be extended for categorical data. In addition, factor scoring methods [180] could be proposed to be able to conduct between and within-patient comparisons. Both the time-varying models and factor analysis are part of thriving fields which undergoes new developments every year regarding new tools and indices to evaluate the models but also cases where data comes from studies where the model assumptions are strongly neglected [66, 181].
Finally, the ultimate goal of a relevant evaluation of patients' pain-related health is to be able to propose therapies that will allow them to improve their health as much as possible, in a multidimensional manner. Thanks to the mixture of longitudinal factor analysers model proposed here, we might be able to extract new scores which can evaluate patients more adequately than the current approaches. Patient health can now be represented by a combination of the different dimensions of pain, specific to the patient according to his or her intrinsic characteristics, history of the disease and current state in their pathway. Afterwards, different therapies could be compared adequately which will allow:

- ✓ Better treatments comparisons in the research literature including randomized controlled trials. Currently, the majority of studies on pain treatments use VAS of

pain intensity as a primary outcome [182]. As discussed before, exclusive pain intensity evaluation does not represent patients global health [183]. In addition to a multidimensional evaluation of pain, our model would focus on evaluating each patient based on how his/her individual outcomes correlate with each other depending on the mixture component he/she belongs to. This would improve the scientific literature on chronic pain treatment by allowing a fair comparison of patient outcomes.

- ✓ To provide physicians with a new evaluation tool which although might take a little more time to assess their patients objectively beyond the usual pain intensity and verbal satisfaction. An evaluation tool with which the patient might feel "understood" and the physician observes an objective improvement or deterioration could potentially improve the patient-physician relationship.
- ✓ To potentially help healthcare organization in their decision making regarding the recommendations and reimbursements of treatments. For example, spinal cord stimulation, which consists of surgically implanting the patient with an expensive medical device that sends electricity at low levels into the spinal cord to relieve pain, is currently implanted in two-steps. The patient is first implanted with a lead placed directly in the spinal column (epidural space) which is connected to an external temporary generator. After the surgery, the patient returns home for 7 days in order to evaluate the benefit of the stimulation. Following the 7-days the patient returns to the hospital in order to be evaluated by the implanting surgeon. For spinal cord stimulation, national and international guidelines recommend permanent implantation of neurostimulation devices based on a 50% pain intensity (VAS) decrease following the 7 days trial period [184, 185]. This recommendation, when followed strictly might prevent patients with less than 50% pain intensity decrease from receiving this treatment, although they might observe an improved quality of life after the trial period. This recommendation could be revisited if adequate multidimensional patient-specific evaluations are proposed.

Adequate patient-specific evaluation of chronic pain would also improve the development in predictive medicine and treatment recommendation systems. Currently, only few authors consider composite outcomes when developing machine learning models to predict whether a patient would or not respond to a given therapy [186, 187] (see our paper which can be found in the annexe 6.2). When trying to predict the patient outcome following a treatment, identifying the appropriate outcome is primordial. Predicting an outcome which does not represent the patient health would just add more noise and complexity to the model and would yield to wrong inferences. Following this thesis, we would also like to develop several new research directions:

1. In the mixture of longitudinal factor analysis model, we only considered time in the mixed effect model evaluating the relationship between the latent factors and time-varying covariates. However, one form of measurement non-invariance comes from the evolution over time of the constructs structure (*e.g.* cultural evolution) and not just from differences between groups. Only latent groups measurement non-invariance was discussed in this thesis. Time-invariance would also be interesting to include in the

model. This would yield to a measurement model of the form:

$$y_{ijt} = \sum_{c=1}^{C} \mathbb{1}_{\{v_i=c\}}(\Lambda_{jtc}\eta_{i.tc} + \mu_j + \epsilon_{ijtc}), \tag{6.1}$$

where λ_{jtc} is the factor loading associated with the outcome j at time t for class c. Contrary to the model discussed in chapter 5, this model allows the factor loadings λ_{jtc} to change over time. One possible approach to achieve this is by considering factor loadings to be polynomial functions of time (*e.g.* $\lambda_{jtc} = \lambda_{jc0} + t\lambda_{jc1} + t^2\lambda_{jc2}$). Another consideration that might be of interest is to allow the latent class variable v to change over time by either writing $v_i = v_i(t)$ where $v_i(t)$ is a smooth function of time estimated using non-parametric modeling, similar to the time-varying mixture proposed in chapter 3 or by writing $v_i = v_{i0} + v_{i1}t + v_{i2}t^2$ as a polynomial function of time. Incorporating time-varying factor loadings and time-varying mixture proportions would allow the model to adjust its structure at different time points of the patient care pathway.

2. After obtaining factor scores, we aim to compare them between the different treatments proposed to patients with chronic pain after spinal surgery of the PREDIBACK study. Cohort observational data where different treatments are proposed to different patients based on their characteristics are generally biased since these characteristics might constitute confounding factors in treatments efficacy. One way to solve the confounding variables problem is to include all these variables in the mixed effect model part of the mixture of longitudinal factor analysis model as covariates. This however is not very efficient when the number of confounding variables is high. Another method used to obtain better treatment effect estimates is the propensity score weighting [188, 189] in which the probability of receiving one of two treatments given the confounding factors is modeled using logistic regression (or other binary classification algorithms depending on the complexity of the relationship between treatment "allocation" and the confounding variables). This probability is then used as a similarity metric to balance patients characteristics (*i.e.* patients who have the same probability of receiving a treatment are matched together). We could incorporate propensity score weighting in the mixture of longitudinal factor analysis models proposed in chapter 5 by using a weighted mixed effect model instead of the classical mixed effects models. In other words, we will weight the observations according to the inverse of their propensity score when estimating the treatment effect. The weights are given by $w_i = T_i P(T_i = 1)/p_i + (1 - T_i)P(T_i = 0)/(1 - p_i)$ where T_i is the treatment binary variable (1 when the patient i is treated and 0 otherwise). the quantity p_i is the probability that patient i will be treated given his characteristics while $P(T_i = 0)$ is the proportion of the treated patients. In order to obtain the treatment effects estimates, the log-likelihood of the mixed effect model component should be multiplied by the weights w_i. Using these methods, we could be able to compare the large panel of treatments proposed to these patients in order to identify treatments which significantly improves the different latent factors for the patients of each latent class.

3. Finally, we would like to use machine learning models to predict patients global health as measured by the factor scores obtained previously, based on the interaction between treatment and patient characteristics. This would potentially help us to develop a treatment recommendation system which would rationalize and simplify the chronic pain patient pathway. This could be achieved by either simply using the structural part of the model proposed in chapter 5 (*i.e.* the mixed effect model describing the relationship between the factors and observed covariates given by: $\eta_{ikct} = X_{ikct}\beta_{kc} + Z_{ikct}\xi_{ikc} + \omega_{ict}$ where the matrices X and Z should contain patients covariates and the treatments proposed to them to assess their impact on patients factor scores in a joint manner. Another approach which might offer more flexibility is to use factor scores extraction methods, which are used to obtain the posterior distribution of the latent variables based on the patients responses to the measured outcomes. These methods include the Bartlet and Thurstone factor scoring. After obtaining the scores of each individual at each time point, a machine learning model such as random forest, support vector machine or neural networks could be used to develop a model to predict a new patient's score given the proposed treatment and his/her characteristics.
4. The work on inferences using hypothesis testing is not proposed in this thesis. However, it is of high importance to be able to conduct different hypothesis test in the proposed models to allow better inferences. For the time-varying mixture of varying mixed effects models proposed in chapter 4, we would like to test the hypothesis that the time-varying coefficients are non-constant time functions. The null and alternative hypothesis of the test are the following: $H_0 : \beta_c(t) = constant_c$ and $H_1 : \beta_c(t) \neq constant_c$. This would allow us to decide whether all the model coefficients should be time-varying (fully non-parametric model) or if some coefficients could be replaced by unknown constants (semi-parametric model). This would reduce the computational burden of the model and facilitate the interpretation of the constant coefficients. A generalized likelihood ratio test has been proposed by Huang *et al.* [85]. The problem with applying a likelihood ratio test in our setting is that we did not have the expression of the observed likelihood function of the model which is needed to find the asymptotic distribution of the test statistic. Huang *et al.* have shown that under some regularity conditions (including the independance of data points which is not verified for our model), the likelihood ratio test has an asymptotic Chi squared distribution with a degree of freedom which depend on the complexity of the kernel function and the number of variables and mixture components. To conclude, in order to be able to develop the discussed hypothesis test, we need to be able to study the asymptotic properties of the difference between the observed data likelihoods under the null and the alternative hypotheses and find the assumptions needed for the test to verify the Wilks phenomenon (*i.e.* the test statistic converges to a Chi-squared distribution with a degree of freedom to identify).

Appendices

Use of Machine learning algorithms to predict the efficacy of spinal cord stimulation in chronic pain after spinal surgery patients and potential applications as an alternative to lead-trial: A predictive multicenter study

Résumé du chapitre en français

La douleur chronique après une chirurgie de la colonne vertébrale peut être traitée avec succès par la stimulation de la moelle épinière (SME). Les directives internationales recommandent fortement qu'un essai de sonde soit effectué avant toute implantation permanente. Des données cliniques récentes mettent en évidence certaines limites majeures de cette approche. Tout d'abord, il semble que les résultats des patients, AVEC ou SANS essai de sonde, soient similaires. En revanche, au cours de l'essai, le taux d'infection chute radicalement dans le temps et peut compromettre le traitement. En nous appuyant sur l'expérience de l'évaluation de la douleur composite et sur des recherches antérieures, nous avons émis l'hypothèse que les modèles d'apprentissage automatique pourraient être des outils de dépistage robustes et des prédicteurs fiables de l'efficacité du SCS à long terme. Nous avons développé plusieurs algorithmes, dont la régression logistique, la régression logistique régularisée (RLR), le classifieur de Bayes naïf, les réseaux de neurones artificiels, la forêt aléatoire et les arbres boostés par le gradient, afin de tester cette hypothèse et d'effectuer des validations internes et externes, l'objectif étant de confronter les prédictions des modèles aux résultats des essais principaux en utilisant un résultat composite à un an de 103 patients. Alors que presque tous les modèles ont démontré leur supériorité sur les essais principaux, le modèle RLR semble représenter le meilleur compromis entre complexité et interprétabilité dans la prédiction de l'efficacité du SCS. Ces résultats soulignent la nécessité d'utiliser la médecine prédictive basée sur l'IA, comme une approche mathématique synergique, visant à aider les implanteurs à optimiser leurs choix cliniques

dans la pratique quotidienne.

Introduction

Chronic Pain After Spinal Surgery (CPASS), formerly called Failed back surgery syndrome (FBSS), is a clinical pathology in which patients present with a set of symptoms encountered after they have had one or more surgical procedures on the lumbar spine for correcting their disc related pathology [37, 190, 191]. The Principal symptom of this pathology is a persistent recurrent pain mainly in the region of the lower back and legs that is generally resistant to physiotherapy and pharmacological treatment [192].
It is estimated that CPASS occurs at a rate ranging from 10% to 40% [193, 194]. This highly frequent pathology generates a severe social, financial and psychological disability for a significant number of patients.
Although chronic pain is often treated using the first-line therapies, the medical community now knows that opioids can be ineffective to treat chronic pain [195] and can induce dramatic consequences on patient's health-related quality of life (*e.g.* traffic incidents, violent behavior, suicidal behavior, addiction, overdose, etc.) [196, 197]. The second opportunity is the use of non-drug therapies. These therapies are numerous, specific and very different. The large panel of non-drug therapies includes physiotherapy, psychotherapy, and transcutaneous electrical nerve stimulation. The surgical therapies include spinal cord stimulation (SCS) which is an invasive neuromodulation treatment modality that has shown a tremendous potential in the management of CPASS.
Several clinical trials and observational studies have been developed in order to assess the utility and cost-effectiveness of SCS for FBSS related leg and/or back pain [198, 199]. Evaluation methods included pain scores, quality of life, functional capacity, and patient satisfaction. Ongoing studies include the prospective, randomized trial comparing SCS associated with conventional medical management with conventional medical management alone in patients with FBSS and predominantly lower back pain. The study aims to compare the outcomes such as pain scores, functional disability, return to work and functional utilization between the two groups. Neuromodulation technological advances including new waveforms and new neural targets are expected to improve the already promising outcomes of neurostimulation techniques for FBSS patients.
Because of the invasive nature of SCS and the lack of large cost-effectiveness studies, SCS is still considered as a last resort therapy. Similarly, to other FBSS treatment modalities, SCS can have divergent outcomes for different patients. It is known that SCS does not relieve certain profiles of patients [200]. Therefore, a two-step procedure for patient selection with a screening trial before the implantation of the SCS device has become a routine. Screening trial starts with the patient being implanted in an invasive surgical procedure with a lead or leads introduced to the epidural space of the spinal cord. The leads are then connected to an external temporary pulse generator. The patient is then sent home in order for him/her to evaluate if SCS generates good pain relief. This trial period lasts from 5 to 10 days.
The screening trial utility consists of determining if the patient is a suitable candidate for SCS by evaluating the pain relief, the paresthesia coverage of the painful region (when conventional SCS is used), the tolerance to this paresthesia and the different stimulation settings and programs. The SCS trial procedure can lead to complication like electrode

migration, dural puncture during electrode placement and/or infection [201].
SCS effects diminish over time in some patients. The extent of decrease of SCS effects over time is reported variously in the literature. Some studies state that the effect has decreased 25%-50% after 2 years [202, 203], but others only see a slight loss of efficacy over time [203, 204]. Therefore, the screening trial predictive power of SCS efficacy also decreases with time. Because of the trial related complications and the decrease of the predictive power of the screening trial, we aim to show in this article that the use of predictive modeling and machine-learning algorithms [205] can yield to a better forecast of SCS long term efficacy compared to the screening trial that is conducted through a surgical procedure that can lead to severe complications.
Our aim in this chapter is to propose some machine learning models that can be used in order to achieve good predictive ability of SCS efficacy and therefore, in the near future, replace the screening trial as a primary SCS patient's selection tool.
In this chapter, we will describe the source of our data and the different predictive factors that we collected from the literature. We will also give a brief explanation of the different statistical and machine learning algorithms used in this article. Thereafter, the predictive power of our models compared to the screening trial will be summarized. Lastly, we will discuss the clinical validation of these machine-learning models and propose the research priorities required for their validation.

Material and methods

Patient data

Data from two different prospective comparative studies were used to conduct this work.
First dataset
The first study is ESTIMET [184], which is a multicenter randomized controlled trial, including 115 PSPS-T2 patients eligible for SCS and implanted with surgical multicolumn SCS paddle lead, in 12 French centers with a 1-year follow-up. The study details are available at https://clinicaltrials.gov/ct2/show/NCT01628237. The primary objective of this study was to compare the efficacy of multicolumn SCS programming to the efficacy of monocolumn SCS programming. As part of the ESTIMET study, all subjects provided informed consent and enrolled following ethical committee approval (CPP-Ouest III) [184]. The study population consisted of PSPS-T2 patients with refractory pain, eligible to SCS according to the French guidelines for SCS selection and implantation. Per these guidelines, an average of 7-day screening trial period was mandatory for all study patients. Patients with a 50% pain decrease, or patients for whom the improvement was clinically important according to a patient-implanter agreement, were implanted with a permanent SCS device at the end of the trial. Among the ESTIMET study patients, those who did not try Transcutaneous Electrical Nerve Stimulation (TENS) were removed from the analysis because TENS efficacy belongs to the set of predictive variables used in the development of the models in this chapter (Figure 1). Finally, ninety-one patients who underwent TENS therapy, completed baseline data, and completed the study 1-year follow-up were included in the analysis. ESTIMET study data was used for the training and internal validation process of our models.

Second dataset
The second study is AIVOC (https://clinicaltrials.gov/ct2/show/NCT02821897), which is a monocentric comparative study, including 15 patients implanted with SCS under general vs awake anesthesia at Poitiers University Hospital (France), with a 1-year follow-up. This study examines the effect of target-controlled intravenous infusion on SCS implantation, lead placement optimization using patient intra-operative feedback and SCS efficacy on back pain coverage. Patients in this study were randomized to either be implanted using general anesthesia during lead implantation or to be implanted using target-controlled intravenous anesthesia with active patient-implanter cooperation during the surgery. Three patients were lost to follow-up. The 12 remaining patients who underwent 12-month follow-up from the AIVOC study were used for external validation. For both the ESTIMET and the AIVOC studies, following verification of inclusion/non-inclusion criteria patients were included and evaluated at baseline, in terms of their sociodemographic, psychological, radiological, and clinical characteristics. One-month after inclusion, all patients were implanted with SCS and underwent a permanent trial: An average of 7-day screening trial period was mandatory for all, per French recommendations of ministry of Health. Patients with a 50% pain decrease, or patients for whom the improvement was clinically important according to a patient-implanter agreement, were implanted with a permanent SCS device at the end of the trial.

Studied variables

Primary outcome

We evaluated SCS efficacy using health-related quality of life (EuroQol with five dimensions and three levels (EQ-5D-3L)) [206], functional disability (Oswestry Disability Index (ODI)) [113] and depression (the Montgomery and Asberg Depression Rating Scale (MADRS) [207]. In order to achieve a holistic evaluation, we used Principal Component Analysis (PCA) with one principal component, including the percentage of global VAS decrease, percentage of EQ-5D increase, percentage of ODI decrease, percentage of MADRS decrease, between baseline and 12-month follow-up. The first principal component was taken as a standardized Global Health Improvement Score (GHIS). Patients were considered as responders if they had a GHIS $>= 0$. Patients who had a negative screening trial or a GHIS < 0 were considered as non-responders. This outcome will be used as a binary dependent variable in our SCS efficacy classification problem. The relationship between our outcome and the standard outcomes used in the SCS literature can be found in Table 1 and 2.

Predictive variables

To avoid comparison bias induced by variable selection based on statistical significance, no primary variable selection was conducted. We used 14 variables that were studied in the SCS literature. They included age [208, 209], sex [210], Depression score [208, 211] measured using MADRS, Body Mass Index (BMI) [212], pain syndromes associated with nervous or somatic lesions (hypoesthesia, brush allodynia) [213], pain increase by movement or by sustained position, TENS efficacy [214], EQ-5D index, back and leg VAS, ODI score, the time interval between the first onset of pain and device implantation (in years) [215, 216]

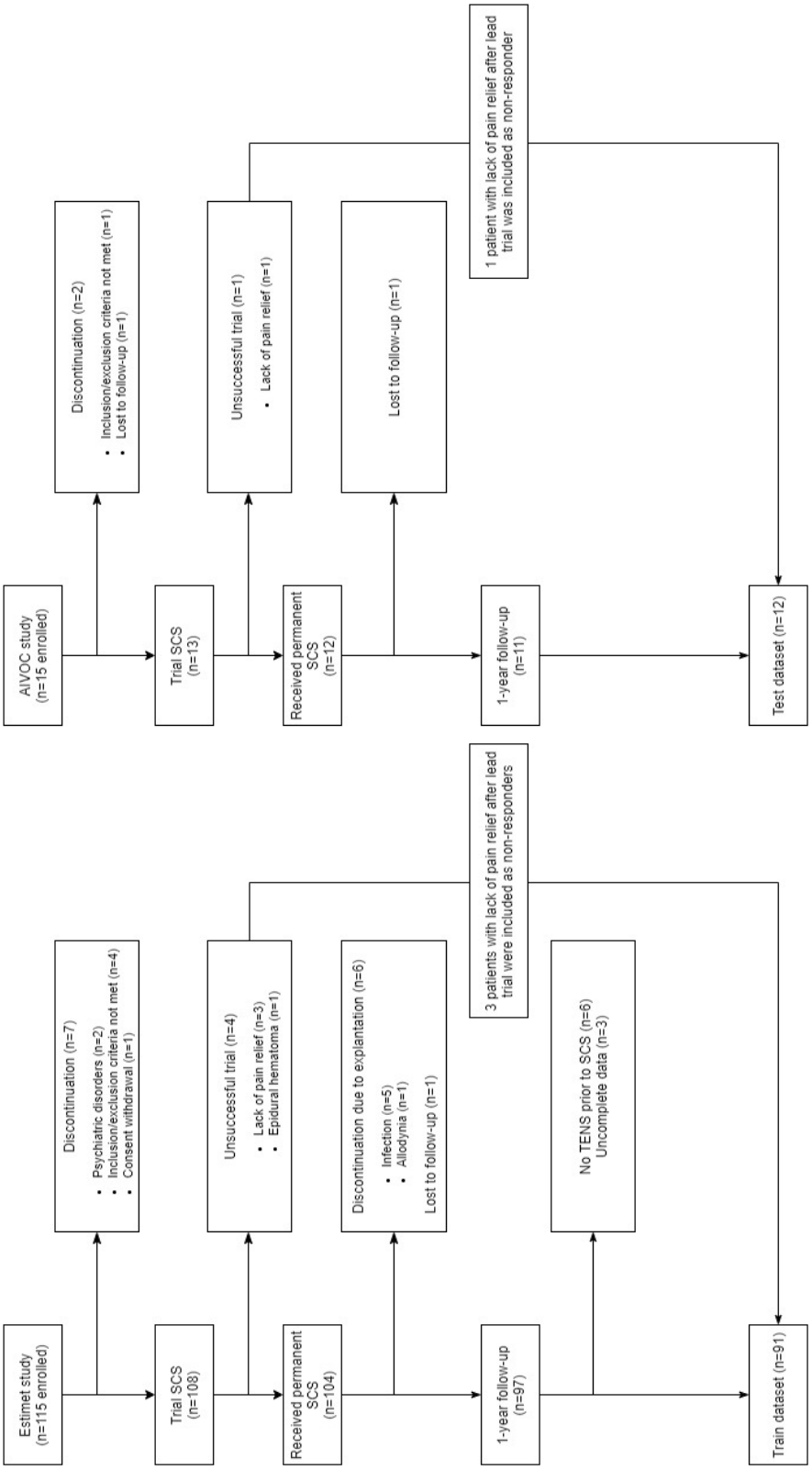

FIGURE 1 – Flowchart of the patients who participated in ESTIMET and AIVOC included in this study.

and the Medication Quantification Scale (MQS III) for measuring medication consumption in chronic pain [186]. Summary of these predictors and the outcome variable can be found in Table 1.

Statistical methods

The statistical analyses were performed using R 3.6.0 software (R Foundation for Statistical Computing, Vienna, Austria).

Descriptive analysis

Categorical variables were described by numbers and percentages, while quantitative variables were described by their means and standard or median standard deviations and interquartile range depending on the skewness of the variable. No missing data imputation was performed.

Multivariate analysis

For this analysis, all data were standardized (subtracting the mean and dividing by the standard deviation) in order to facilitate interpretation and convergence of the iterative models.
In this section, each model and its implementation will be described briefly. The following classification models will be used in this chapter to predict SCS outcome at 12-month follow-up:
Logistic Regression (LR): The logistic regression generalized linear model was trained using the `glm` function available in `R stats` package. Even if logistic regression can detect only linear relations between variables, it is still widely used because of its simplicity and interpretability and it has shown a better performance on simple classification problems where classes can be separated linearly. Backward-forward (bidirectional) stepwise variable selection procedure was used in order to identify the best LR model based on the AIC selection criterion. To avoid overfitting, no interaction terms were included in the model.
Regularized Logistic Regression (RLR): RLR model [217] was developed using the `glmnet` package. Regularization is a technique used to shrink or reduce insignificant effects in the logistic regression to zero. This technique allows the model to avoid overfitting since it reduces model variance. Regularization can also be considered as a variable selection technique. As our data contains very few variables, we opted for the use of ridge regularization. The optimal regularization parameter λ was identified using the cross validation described below.
Support Vector Machine (SVM): SVM models are known for their classification capability, since SVM algorithms are computationally stable, and generalize well giving enough training examples [218]. Beside linear relations, SVM can also detect nonlinearities by transforming input data into a space of higher dimensionality with the use of a kernel function. In this article, we used the radial basis function kernel to allow non-linearity. SVM was developed using the `svm` function available in the `e1071 R` package [219]. The optimal cost and gamma parameters were identified using cross-validation.

Naive Bayes (NB): NB classifier [220] was developed with default hyperparameters using the `naiveBayes` function available in the `e1071` R package.
Artificial Neural Network (ANN): ANN model [221] was developed using the `keras` package [222], which is a popular deep learning `Python` package that has been added recently to R software as a package available in `CRAN`. To avoid overfitting, we used a relatively small ANN. the ANN model contained two hidden layers with eight nodes each. To ensure model convergence, we trained the ANN model. Sigmoid activation functions were used to allow non-linearity. The weights are estimated using the Backpropagation method, which is a gradient-based optimization method, which allows the estimation of the error at the output of the hidden layers neurons, thus enabling the update of weights in the hidden layers using error gradients.
Classification and regression tree (CART): CART Classification tree model was developed using the `rpart` function available in the package with the same name [223]. At first, all patients belong to a simple node representing a responders and non-responders rates. Afterwards the node is split creating two new child nodes. The splitting is done by choosing the predictor and the optimal split point (*e.g.* age>45) that differentiates responders from non-responders. The algorithm stops when the observations inside the nodes are homogenous and further splits are undesirable. The minimum number of observation in a node was set to five.
Random Forest (RF): RF model [224] was trained using the `randomForest` function available in the R package with the same name [225].
Gradient Boosted Trees (GBT): GBT model [226] was developed using the `xgboost` package [227].
Both RF and GBT are tree based ensemble models. Each of this two models use a different ensemble learning technique. RF uses Bagging (Bootstrap aggregating) which can be described as follows: each iteration, a decision tree model is created using data from a bootstrap sample drawn from the training set, independently from other iterations. After growing all the trees, each tree casts a unit vote for the outcome (responder or non-responder) of a new giving data sample. The final prediction for this data sample is the average of predictions obtained from all the trees.
GBT, as their name suggests, use a technique called Boosting, whereby weighted combinations of decision trees are constructed into a stronger classifier in an iterative way (contrary to random forests where the weights are uniform and the trees are grown independently). The strongest classification tree is weighted to count more substantially in the prediction outcome. A tree that most accurately classifies examples that were misclassified by the first tree is grown next. This procedure allows the trees that are weak on some examples to be compensated by a tree that performs better on those same examples.
The optimal hyperparameters of both RF and GBT models were estimated using cross-validation (CV).

Testing data and model assessment

Cross-validation was used to identify the optimal hyperparameters of the models and to assess their internal validity. Our cross validation procedure goes as follows: We divide our training

dataset into 10 separate datasets (10−fold cross-validation). One subset is kept for model assessment. 9−fold cross-validation is conducted on the remaining nine subsets to identify the optimal hyperparameters. The optimal hyperparameters are then used to develop the models. The final models are then tested on the 10^{th} subset. This process is conducted for all the 10 folds. This procedure allows us to reduce evaluation bias associated with identifying the optimal hyperparameters and evaluating the models on the same subset.
The models were evaluated using sensitivity, specificity, accuracy, and Area Under ROC Curve (AUC). Their means and standard deviations are reported.

External validation

An independent data set of the 12 patients from the AIVOC study was used for model assessment and external validation.

Results

Descriptive analysis

Descriptive statistics of our training and testing data can be found in Table 1. The majority of our predictors were homogenous between the training and testing datasets. 49.5% of the training dataset were male and 50.5% were female while 41.7% of the testing dataset were male and 58.3% were female. The mean age of our training sample is 47.7 (9.5) and the mean age the testing sample is 49.5 (14.7). We observed some differences between the training and testing datasets. Patients in the training dataset were less likely to respond to TENS therapy than patients in the testing dataset (52.7% for training dataset vs 83.3% for testing dataset). Pain medication consumption was also higher in training dataset (MQS of 24.5 (14.7) for training dataset vs 5.6 (7.8) for testing dataset).
The results of our PCA leading to the development of our holistic outcome can be found in Table 2. The ODI percentage of decrease had the highest weight in our PCA (0.86) followed by VAS (0.81), MADRS (0.59) and EQ-5D (0.51). In our training dataset, 45 patients (49.5%) had a positive holistic response to SCS and 46 (50.5%) had a negative holistic outcome. Similarly, 6 patients (50%) had a positive outcome and 6 patients (50%) had a negative outcome in the testing dataset. Table 3 shows the relationship between our composite outcome and the classical pain assessment outcomes used in the literature.

Training data results (internal validation)

We developed our models using the eight different binary classification methods, described in the statistical methods section. The SVM model showed the highest performance metrics according to our cross validation procedure results (AUC = 0.801, sd = 0.202). It had a specificity of 81.3% (sd = 14.8%) and a sensitivity of 80.7% (sd = 20.1%). The GBT model also showed acceptable performances with lower variability between folds, which indicate a more stable model (AUC = 0.790 (0.105), Specificity = 80.0% (3.6%), Sensitivity = 70.1% (16.7%)). The two logistic regression models LR and RLR, showed comparable results to the previous models. The regularized logistic regression model had an AUC of 0.781 (0.120),

Variable	Train set descriptive statistics	Test set descriptive statistics
Response variable		
Good composite outcome	45 (49.5%)	6 (50.0%)
Bad composite outcome	46 (50.5%)	6 (50.0%)
Predictors at baseline		
Age	47.7 (9.5)	49.5 (14.7)
Sex		
Male	45 (49.5%)	5 (41.7%)
Female	46 (50.5%)	7 (58.3%)
BMI	27.4 (5.04)	24.6 (3.9)
Pain duration	12.2 (10.7)	15.8 (15.1)
ODI	50.5 (9.1)	44.7 (12.8)
MADRS	16.9 (10.4)	11.3 (8.3)
EQ-5D	0.38 (0.20)	0.54 (0.19)
EQ-5D VAS	45.8 (17.3)	51.1 (20.0)
Leg VAS	75.0 (11.3)	72.9 (16.0)
Back VAS	71.2 (15.1)	67.6 (21.8)
TENS efficacy		
Effective	48 (52.7%)	10 (83.3%)
Not effective	43 (47.3%)	2 (16.7%)
Hypoesthesia		
Yes	28 (30.8%)	2 (16.7%)
No	63 (69.2%)	10 (83.3%)
Allodynia		
Yes	22 (24.2%)	5 (41.7%)
No	69 (75.8%)	7 (58.3%)
Positional pain changes		
Yes	74 (81.3%)	9 (75.0%)
No	17 (18.7%)	3 (25%)
MQS	24.5 (14.7)	5.6 (7.8)

TABLE 1 – Descriptive statistics of our 12-month outcome and baseline predictive variables for both the training and testing datasets.
BMI: Body Mass Index ; ODI: Oswestry Disability Index ; MADRS: Montgomery-Asberg Depression Rating Scale ; EQ-5D: EuroQol-5 Dimensions ; VAS: Visual Analogic Scale ; TENS: Transcutaneous Electrical Nerve Stimulation ; MQS: Medication Quantification Scale.

a sensitivity of 69.8% (12.0%) and a specificity of 73.0% (12.9%). Our logistic regression model had an AUC of 0.779 (0.114), a sensitivity of 70.9% (18.0%) and a specificity of 72.8% (13.3%). The AUCs of the RF, naïve Bayes and CART tree were 0.755 (0.123), 0.697 (0.153) and 0.657 (0.136) respectively. According to these results, we would recommend GBT, SVM or RLR models. The advantage of the RLR model is that it can be interpreted as a simple logistic regression model. The AUC of the lead-trial was 0.670 with a sensitivity of 79.5% and a specificity of 52.4%. Lead-trial specificity is very low. A high percentage of patients

Variables changes (%) between baseline and 12 months	1st principal component loadings (50.1% of the total variance)
ODI	0.86
VAS	0.81
MADRS	0.59
EQ-5D score	0.51

Table 2 – Composition of the first principal component of the PCA of our outcomes.
ODI: Oswestry Disability Index ; MADRS: Montgomery-Asberg Depression Rating Scale ; EQ-5D: EuroQol-5 Dimensions ; VAS: Visual Analogic Scale.

with a bad 1-year outcome had a 50% pain decrease following lead-trial.

Outcomes	Good composite outcome (GHIS >= 0)	Bad composite outcome (GHIS <0)
50% global VAS decrease		
Yes	43 (93.5%)	8
No	3	37 (82.2%)
30% ODI decrease		
Yes	34 (73.9%)	8
No	12	37 (82.2%)
0.19 points change in EQ-5D		
Yes	30 (65.2%)	17
No	16	28 (62.2%)

Table 3 – Relationship between GHIS outcome and VAS decrease, ODI decrease and improvement in EQ-5D.
EQ-5D: EuroQol-5 Dimensions ; ODI: Oswestry Disability Index ; GHIS: Global Health Improvement score ; VAS: Visual Analogue Scale.

External validation

The results of our models using the external validation set can be found in Table 4. The results we obtained on the external validation set were similar to the external set. The best performance was achieved using the RF model, the GBT model and the RLR model. While the RF model showed a lower performance on the training set, the GBT and RLR model had good results on both the training and testing sets. Similarly, to the training set, the testing set lead-trial results showed a good sensitivity (100%) but a bad specificity (33.3%).

Models interpretability

The majority of models discussed and analyzed in this chapter are black box models, meaning that it is barely impossible to extract useful information on how the variables interact with the SCS outcome. However, this interpretability is sometimes disregarded in order to achieve more complexity. In this section, we will show the role of the explicative

Model	True good outcome	True bad outcome
Screening-trial (AUC = 0.69)		
Good outcome	6 (Sensitivity = 100%)	4
Bad outcome	0	2 (Specificity = 33.3%)
LR (AUC = 0.72)		
Predicted good outcome	5 (Sensitivity = 83.3%)	2
Predicted bad outcome	1	4 (Specificity = 66.7%)
RLR (AUC = 0.81)		
Predicted good outcome	5 (Sensitivity = 83.3%)	2
Predicted bad outcome	1	4 (Specificity = 66.7%)
SVM (AUC = 0.75)		
Predicted good outcome	6 (Sensitivity = 100%)	2
Predicted bad outcome	0	4 (Specificity = 66.7%)
NB (AUC = 0.81)		
Predicted good outcome	5 (Sensitivity = 83.3%)	1
Predicted bad outcome	1	5 (Specificity = 83.3%)
ANN (AUC = 0.72)		
Predicted good outcome	5 (Sensitivity = 83.3%)	2
Predicted bad outcome	1	4 (Specificity = 66.7%)
CART (AUC = 0.72)		
Predicted good outcome	4 (Sensitivity = 66.7%)	1
Predicted bad outcome	2	5 (Specificity = 83.3%)
RF (AUC = 0.83)		
Predicted good outcome	5 (Sensitivity = 83.3%)	1
Predicted bad outcome	1	5 (Specificity = 83.3%)
GBT (AUC = 0.81)		
Predicted good outcome	5 (Sensitivity = 83.3%)	1
Predicted bad outcome	1	5 (Specificity = 83.3%)

TABLE 4 – AUC, specificity and sensitivity of screening-trial and our model on the external validation set.

variables in the decision making for the logistic regression model. Table 5 shows the unstandardized coefficients, standardized coefficients and their 95% confidence intervals, for each variable.

The logistic regression model has the advantage of having a straightforward interpretation. Based on the logistic regression model results (Table 5), the baseline psychological state had a significant effect on the outcome of the treatment (MADRS odds ratio = 0.908, p-value = 0.002). Depressive patients have a reduced chance of having a good outcome 1 year after the implantation of the SCS device. We observed a greater probability of achieving a successful outcome in patients with hypoesthesia related to back pain (odds ratio = 10.59, p = 0.0008) and positional back pain symptoms (odds ratio = 4.48, p = 0.043). Patients who had did not achieve a 50% pain decrease after TENS therapy before SCS had lower chances of a successful SCS therapy at 1-year follow-up (odds ratio = 0.269,

p value=0.016).
This logistic regression model can be written as:

$$P(Y = good|X) = \frac{1}{1 + \exp(-(\beta_0 + \sum_{i=1}^{8} \beta_i X_i))}. \quad (2)$$

Good composite outcome Variables (GHIS >= 0)	Unstandardized coefficients (β)	Standardized coefficients	95% CI	p-value
Intercept	-3.044	-0.070	[-0.586 ;0.447]	0.792
Duration of pain	-0.038	-0.041	[-0.939 ;0.116]	0.137
MADRS	-0.097	-1.012	[-1.653 ;-0.371]	0.002**
EQ5D VAS	0.032	0.554	[0.015 ;1.093]	0.044*
Leg VAS	0.040	0.449	[-0.09 ;0.988]	0.102
Hypoesthesia : Yes	2.361	1.096	[0.455 ;1.737]	0.0008***
TENS : Not effective	-1.312	-0.659	[-1.196 ;-0.122]	0.016*
MQS	-0.024	-0.449	[-1.061 ;0.163]	0.151
Positional pain changes : Yes	1.500	0.588	[0.017 ;1.159]	0.043*

TABLE 5 – Standardized coefficients of selected variables, confidence intervals and significance levels.
MADRS: Montgomery-Asberg Depression Rating Scale ; EQ-5D: EuroQol-5 Dimensions ; VAS: Visual Analogic Scale ; TENS: Transcutaneous Electrical Nerve Stimulation ; MQS: Medication Quantification Scale.

Discussion

In this chapter, we used data from two RCT studies to investigate retrospectively the performance of machine-learning models for predicting long-term SCS outcome. Our results show that machine-learning techniques can lead to a more accurate prediction of the patient's SCS long-term outcome than the lead screening trial. A large amount of studies tried to identify the predictors of SCS outcome but very few used these predictors in a multivariate statistical model in order to propose a model that can be used for predicting the efficacy of SCS for FBSS patients in daily practice [186]. We also proposed a new holistic approach to evaluate patients' well-being based on PCA conducted on pain intensity, functional disability, quality of life and psychological distress measures. We showed that our holistic pain evaluation was significantly associated with these four pain dimensions. 49.5% of our sample were considered as responders while 50.5% were considered as non-responders.
We accurately predicted the efficacy of SCS better than the lead-trial (AUC=0.69 for lead trial vs AUC=0.81 for the RLR model). We recommend the RLR model as it achieved an acceptable performance in both internal and external validation. The RLR model is also easily interpretable (It can be interpreted like a logistic regression model) and variable selection can be conducted in an automated manner which can simplify the creation of medical decision support tools.

Few authors proposed models for predicting SCS efficacy. The most recent paper on the subject is the paper by Goudman et al. [186]. In their paper, they developed a logistic regression model for predicting high density SCS efficacy using data from 92 FBSS patients and a set of variables including age, sex, back and leg pain intensity, MQS, ODI, Pittsburg sleep quality index, EQ-5D and second order interactions between these variables. They achieved a 90% specificity and sensitivity on an out of sample dataset consisting of 20% of their dataset. Cross-validation results and lead trial predictive value were not reported. Their model showed some non-diagnosed problems. Their model contains some coefficients that are very large compared to the scale of the variable. This might be due to the inclusion of a large amount of interaction terms for a moderate sample size. Although their model had a good performance, our model seems more stable. Sparkes et al. [228] also proposed a logistic regression model for predicting pain decrease at 12-month follow-up using data from 56 FBSS patients. Their model included age, sex, duration of pain, anxiety and depression scores and coping strategies. No out-of-sample validation was conducted in their study. Our logistic regression model shows that patients with higher levels of depressive symptoms are less likely to benefit from SCS. Patients with a higher perceived health status (EQ-5D VAS) were more likely to achieve a good outcome following SCS. TENS efficacy which is a well-studied predictor of SCS efficacy, was also significantly associated with SCS efficacy. Lastly, presence of back pain hypoesthesia and changes in pain depending on the patient's position were associated with a greater likelihood of a good outcome following SCS. These results suggest that psychological evaluation and pain typology are important in patient's selection prior to SCS implantation.
Apart from the relatively large sample size, the strengths of this study include having paid the most careful attention to the use of a large variety of models while reducing the biases associated with this type of research to assure a maximal generalizability of our models. The multicenter nature of our sample also helps in ensuring generalizability and applicability to clinical practice. The development of a composite outcome that is based on objective methods may also result and better patients' well-being evaluation and therefore a more precise forecast of patients' outcome.
Despite these strengths, our study is not without limitations. Thereafter, we will discuss the points that need to be treated and some research idea that might help addressing those points.

Paresthesia intolerance

One of the reasons a patient might fail the SCS trial is the inability to tolerate SCS induced paresthesia. SCS induced paresthesia intolerance is not taken into consideration in our outcomes. It has been shown that there is a high correlation between the intolerance to SCS and TENS induced paresthesia [214]. Therefore, a combination of a TENS trial and our machine-learning model can serve as a non-invasive screening tool prior to SCS permanent implantation. Furthermore, with the development of new paresthesia-free neurostimulation waveforms like high frequency and BURST stimulation, paresthesia intolerance as an SCS trial criterion will be obsolete.

Models with small sample size

One of the drawbacks of our study is the small sample size, which we tried to overcome by performing variable selection based on literature instead of an exploratory "significance level" selection. Larger scale studies are needed in order to obtain more robustness and a better validation of our models.

Further evaluation methodology

Although we have shown that our models predictive power is superior to the lead-trial, it is still necessary to conduct a more robust comparative study to assess the overall superiority of utilizing machine-learning models instead of lead-trial. The evaluation criteria of this study should cover the predictive ability of the screening trial and mathematical models, medico-economical evaluation and surgery related adverse events. A study such as described previously can lead to a complete replacement of the SCS trial procedure by a more objective, non-invasive and accurate procedure based on a TENS trial and machine-learning models which reduces the risks of complications associated with invasive surgical procedures by selecting more efficiently the patients who may benefit from spinal cord stimulation.

Conclusion

Machine-learning and statistical models are gaining a lot of attention lately but their application to SCS efficacy data is still limited. We showed that it might be of great interest to adopt such methods for SCS data in order to achieve a more precise, cost-effective and safe tools patient selection for SCS implantation. Therefore, further investigations and large-scale studies need to be conducted.

Bibliographie

[1] M. B. Toledano, R. B. Smith, J. P. Brook, M. Douglass, and P. Elliott. How to establish and follow up a large prospective cohort study in the 21st century–lessons from uk cosmos. *PloS one*, 10(7):e0131521–e0131521, July 2015. ISSN 1932-6203. doi: 10.1371/journal.pone.0131521. URL https://pubmed.ncbi.nlm.nih.gov/26147611.

[2] A. M. D. Brisco and S. Migliorati. A new mixed-effects mixture model for constrained longitudinal data. *Statistics in medicine*, 2019. doi: https://doi.org/10.1002/sim.8406.

[3] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38(4):963–974, 1982. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2529876.

[4] K. Meyer. Maximum likelihood estimation of variance components for a multivariate mixed model with equal design matrices. *Biometrics*, 41(1):153–165, 1985. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2530651.

[5] C. Proust-Lima, H. Amièva, and H. Jacqmin-Gadda. Analysis of multivariate mixed longitudinal data: a flexible latent process approach. *The British journal of mathematical and statistical psychology*, 66 3:470–87, 2013.

[6] J. Roy and X. Lin. Latent variable models for longitudinal data with multiple continuous outcomes. *Biometrics*, 56(4):1047–1054, 2000. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/2677036.

[7] X. An, Q. Yang, and P. M. Bentler. A latent factor linear mixed model for high-dimensional longitudinal data analysis. *Statistics in medicine*, 32(24):4229–4239, October 2013. ISSN 0277-6715. URL https://pubmed.ncbi.nlm.nih.gov/23640746.

[8] J. J. McArdle. *Dynamic but structural equation modeling of repeated measures data.* Perspectives on individual differences. Plenum Press, New York, NY, US, 1988. doi: 10.1007/978-1-4613-0893-5_17.

[9] M. W. Browne and J. R. Nesselroade. *Representing Psychological Processes With Dynamic Factor Models: Some Promising Uses and Extensions of Autoregressive Moving Average Time Series Models.* Multivariate applications book series. Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 2005.

[10] C. Proust, H. Jacqmin-Gadda, J. M. G. Taylor, J. Ganiayre, and D. Commenges. A nonlinear model with latent process for cognitive evolution using multivariate longitudinal data. *Biometrics*, 62(4):1014–1024, December 2006. ISSN 0006-341X. doi: 10.1111/j.1541-0420.2006.00573.x. URL https://pubmed.ncbi.nlm.nih.gov/17156275.

[11] K. V. Flannery. The cultural evolution of civilizations. *Annual Review of Ecology and Systematics*, 3:399–426, 1972. ISSN 00664162. URL http://www.jstor.org/stable/2096853.

[12] G. McLachlan, D. Peel, and R. Bean. Modelling high-dimensional data by mixtures of factor analyzers. *Computational Statistics & Data Analysis*, 41(3):379–388, 2003. ISSN 0167-9473. doi: https://doi.org/10.1016/S0167-9473(02)00183-4. URL https://www.sciencedirect.com/science/article/pii/S0167947302001834. Recent Developments in Mixture Model.

[13] C. Chenaf, J. Delorme, N. Delage, D. Ardid, A. Eschalier, and N. Authier. Prevalence of chronic pain with or without neuropathic characteristics in france using the capture-recapture method: a population-based study. *Pain*, 159(11):2394—2402, November 2018. ISSN 0304-3959. doi: 10.1097/j.pain.0000000000001347. URL https://doi.org/10.1097/j.pain.0000000000001347.

[14] R.-D. Treede, W. Rief, A. Barke, Q. Aziz, M. I. Bennett, R. Benoliel, M. Cohen, S. Evers, N. B. Finnerup, M. B. First, M. A. Giamberardino, S. Kaasa, E. Kosek, P. Lavand'homme, M. Nicholas, S. Perrot, J. Scholz, S. Schug, B. H. Smith, P. Svensson, J. W. S. Vlaeyen, and S.-J. Wang. A classification of chronic pain for icd-11. *Pain*, 156 (6):1003–1007, June 2015. ISSN 0304-3959. URL https://pubmed.ncbi.nlm.nih.gov/25844555.

[15] S. F. d'Étude et de Traitement de la Douleur (SFETD). Livre blanc de la douleur 2017 État des lieux et propositions pour un système de santé éthique, moderne et citoyen. https://www.sfetd-douleur.org/wp-content/uploads/2019/09/livre_blanc-2017-10-24.pdf, 2017.

[16] K. Hider-Mlynarz, P. Cavalié, and P. Maison. Trends in analgesic consumption in france over the last 10 years and comparison of patterns across europe. *British journal of clinical pharmacology*, 84(6):1324–1334, June 2018. ISSN 0306-5251. URL https://pubmed.ncbi.nlm.nih.gov/29514410.

[17] C. Chenaf, J.-L. Kaboré, J. Delorme, B. Pereira, A. Mulliez, M. Zenut, N. Delage, D. Ardid, A. Eschalier, and N. Authier. Prescription opioid analgesic use in france: Trends and impact on morbidity–mortality. *European Journal of Pain*, 23(1):124–134, 2019. doi: https://doi.org/10.1002/ejp.1291. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ejp.1291.

[18] M. Dueñas, B. Ojeda, A. Salazar, J. A. Mico, and I. Failde. A review of chronic pain impact on patients, their social environment and the health care system. *Journal of Pain Research*, 9:457–467, June 2016. ISSN 1178-7090. doi: 10.2147/JPR.S105892. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4935027/.

[19] J. C. Ballantyne and M. D. Sullivan. Intensity of chronic pain — the wrong metric? *New England Journal of Medicine*, 373(22):2098–2099, 2015. doi: 10.1056/NEJMp1507136. URL https://doi.org/10.1056/NEJMp1507136. PMID: 26605926.

[20] J. A. Hashmi, M. N. Baliki, L. Huang, A. T. Baria, S. Torbey, K. M. Hermann, T. J. Schnitzer, and A. V. Apkarian. Shape shifting pain: chronification of back pain shifts brain representation from nociceptive to emotional circuits. *Brain : a journal of neurology*, 136(Pt 9):2751–2768, September 2013. ISSN 0006-8950. URL https://pubmed.ncbi.nlm.nih.gov/23983029.

[21] F. M. Kovacs, J. Seco, A. Royuela, A. Peña, A. Muriel, and Spanish Back Pain Research Network. The correlation between pain, catastrophizing, and disability in subacute and chronic low back pain: a study in the routine clinical practice of the Spanish National Health Service. *Spine*, 36(4):339–345, February 2011. ISSN 1528-1159. doi: 10.1097/BRS.0b013e3181cfba29.

[22] S. L. Scholich, D. Hallner, R. H. Wittenberg, M. I. Hasenbring, and A. C. Rusu. The relationship between pain, disability, quality of life and cognitive-behavioural factors in chronic back pain. *Disability and Rehabilitation*, 34(23):1993–2000, 2012. ISSN 1464-5165. doi: 10.3109/09638288.2012.667187.

[23] M. D. Sullivan and J. C. Ballantyne. Must we reduce pain intensity to treat chronic pain? *PAIN*, 157(1), 2016. ISSN 0304-3959. URL https://journals.lww.com/pain/Fulltext/2016/01000/Must_we_reduce_pain_intensity_to_treat_chronic.7.aspx.

[24] M. A. Hadi, G. A. McHugh, and S. J. Closs. Impact of chronic pain on patients' quality of life: A comparative mixed-methods study. *Journal of patient experience*, 6(2):133–141, June 2019. ISSN 2374-3743. URL https://pubmed.ncbi.nlm.nih.gov/31218259.

[25] J. Eriksen, P. Sjøgren, E. Bruera, O. Ekholm, and N. K. Rasmussen. Critical issues on opioids in chronic non-cancer pain:. *PAIN*, 125(1), 2006. ISSN 0304-3959. URL https://journals.lww.com/pain/Fulltext/2006/11000/Critical_issues_on_opioids_in_chronic_non_cancer.21.aspx.

[26] C. L. Maroney, A. Litke, D. Fischberg, C. Moore, and R. S. Morrison. Acceptability of severe pain among hospitalized adults. *Journal of palliative medicine*, 7(3):443—450, June 2004. ISSN 1096-6218. doi: 10.1089/1096621041349563. URL https://doi.org/10.1089/1096621041349563.

[27] M. C. Bushnell, M. Ceko, and L. A. Low. Cognitive and emotional control of pain and its disruption in chronic pain. *Nature reviews. Neuroscience*, 14(7):502–511, July 2013. ISSN 1471-003X. URL https://pubmed.ncbi.nlm.nih.gov/23719569.

[28] P. Rigoard, A. Ounajim, L. Goudman, P.-Y. Louis, Y. Slaoui, M. Roulaud, N. Naiditch, B. Bouche, P. Page, B. Lorgeoux, S. Baron, E. Charrier, L. Poupin, D. Rannou, G. B. de Montgazon, B. Roy-Moreau, N. Grimaud, N. Adjali, K. Nivole, M. Many, R. David, C. Wood, R. Rigoard, M. Moens, and M. Billot. A novel multi-dimensional clinical response index dedicated to improving global assessment of pain in patients with

persistent spinal pain syndrome after spinal surgery, based on a real-life prospective multicentric study (prediback) and machine learning techniques. *Journal of Clinical Medicine*, 10(21), 2021. ISSN 2077-0383. doi: 10.3390/jcm10214910. URL https://www.mdpi.com/2077-0383/10/21/4910.

[29] H. Breivika. Fifty years on the visual analogue scale (vas) for pain-intensity is still good for acute pain. but multidimensional assessment is needed for chronic pain. *Scandinavian Journal of Pain*, 11(1):150–152, 2016. doi: doi:10.1016/j.sjpain.2016.02.004. URL https://doi.org/10.1016/j.sjpain.2016.02.004.

[30] D. Dorfman, M. C. George, J. Robinson-Papp, T. Rahman, R. Tamler, and D. M. Simpson. Patient reported outcome measures of pain intensity: Do they tell us what we need to know ? *Scandinavian Journal of Pain*, 11(1):73–76, 2016. doi: doi:10.1016/j.sjpain.2015.12.004. URL https://doi.org/10.1016/j.sjpain.2015.12.004.

[31] K. Talbot, V. J. Madden, S. L. Jones, and G. L. Moseley. The sensory and affective components of pain: are they differentially modifiable dimensions or inseparable aspects of a unitary experience ? a systematic review. *British journal of anaesthesia*, 123 (2):263–272, August 2019. ISSN 0007-0912. URL https://pubmed.ncbi.nlm.nih.gov/31053232.

[32] L. Litcher-Kelly, S. A. Martino, J. E. Broderick, and A. A. Stone. A systematic review of measures used to assess chronic musculoskeletal pain in clinical and randomized controlled clinical trials. *The journal of pain*, 8(12):906–913, December 2007. ISSN 1528-8447. URL https://pubmed.ncbi.nlm.nih.gov/17690014.

[33] R. H. Dworkin, D. C. Turk, J. T. Farrar, J. A. Haythornthwaite, M. P. Jensen, N. P. Katz, R. D. Kerns, G. Stucki, R. R. Allen, N. Bellamy, D. B. Carr, J. Chandler, P. Cowan, R. Dionne, B. S. Galer, S. Hertz, A. R. Jadad, L. D. Kramer, D. C. Manning, S. Martin, C. G. McCormick, M. P. McDermott, P. McGrath, S. Quessy, B. A. Rappaport, W. Robbins, J. P. Robinson, M. Rothman, M. A. Royal, L. Simon, J. W. Stauffer, W. Stein, J. Tollett, J. Wernicke, and J. Witter. Core outcome measures for chronic pain clinical trials: Immpact recommendations. *PAIN*, 113(1), 2005. ISSN 0304-3959. URL https://journals.lww.com/pain/Fulltext/2005/01000/Core_outcome_measures_for_chronic_pain_clinical.5.aspx.

[34] E. J. Cassel. The nature of suffering and the goals of medicine. *New England Journal of Medicine*, 306(11):639–645, 1982. doi: 10.1056/NEJM198203183061104. URL https://doi.org/10.1056/NEJM198203183061104. PMID: 7057823.

[35] M. Kyranou and K. Puntillo. The transition from acute to chronic pain: might intensive care unit patients be at risk ? *Annals of intensive care*, 2(1):36–36, August 2012. ISSN 2110-5820. URL https://pubmed.ncbi.nlm.nih.gov/22898192.

[36] D. Borsook, A. M. Youssef, L. Simons, I. Elman, and C. Eccleston. When pain gets stuck: the evolution of pain chronification and treatment resistance. *Pain*, 159(12): 2421–2436, December 2018. ISSN 0304-3959. URL https://pubmed.ncbi.nlm.nih.gov/30234696.

[37] S. A. Schug, P. Lavand'homme, A. Barke, B. Korwisi, W. Rief, R.-D. Treede, and IASP Taskforce for the Classification of Chronic Pain. The IASP classification of chronic pain for ICD-11: chronic postsurgical or posttraumatic pain. *Pain*, 160(1):45–52, January 2019. ISSN 1872-6623. doi: 10.1097/j.pain.0000000000001413.

[38] P. Rigoard, K. Gatzinsky, J.-P. Deneuville, W. Duyvendak, N. Naiditch, J.-P. Van Buyten, and S. Eldabe. Optimizing the Management and Outcomes of Failed Back Surgery Syndrome: A Consensus Statement on Definition and Outlines for Patient Assessment. *Pain Research & Management*, 2019:3126464, February 2019. ISSN 1918-1523. doi: 10.1155/2019/3126464. URL https://europepmc.org/article/pmc/pmc6398030.

[39] T. Hastie and R. Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55(4):757–796, 1993. ISSN 00359246. doi: https://doi.org/10.1111/j.2517-6161.1993.tb01939.x. URL http://www.jstor.org/stable/2345993.

[40] R. D. De Veaux. Mixtures of linear regressions. *Computational Statistics & Data Analysis*, 8(3):227–245, November 1989. doi: https://doi.org/10.1016/0167-9473(89)90043-1. URL https://ideas.repec.org/a/eee/csdana/v8y1989i3p227-245.html.

[41] A. Ounajim, M. Billot, P.-Y. Louis, Y. Slaoui, D. Frasca, L. Goudman, M. Roulaud, N. Naiditch, B. Lorgeoux, S. Baron, K. Nivole, M. Many, N. Adjali, P. Page, B. Bouche, E. Charrier, L. Poupin, D. Rannou, G. B. de Montgazon, B. Roy-Moreau, C. Wood, R. Rigoard, R. David, M. Moens, and P. Rigoard. Finite mixture models based on pain intensity, functional disability and psychological distress composite assessment allow identification of two distinct classes of persistent spinal pain syndrome after surgery patients related to their quality of life. *Journal of clinical medicine*, 10(20):4676, October 2021. ISSN 2077-0383. URL https://pubmed.ncbi.nlm.nih.gov/34682799.

[42] G. K. Robinson. That blup is a good thing: The estimation of random effects. *Statist Sci*, 6(1):15–32, 02 1991. doi: 10.1214/ss/1177011926. URL https://doi.org/10.1214/ss/1177011926.

[43] T. P. Garcia and Y. Ma. Optimal estimator for logistic model with distribution-free random intercept. *Scandinavian Journal of Statistics*, 43:156–171, 2016.

[44] C. McCulloch and J. Neuhaus. Misspecifying the shape of a random effects distribution: Why getting it wrong may not matter. *Statistical Science*, 26:388–402, 2011.

[45] M. Delattre, M. Lavielle, and M.-A. Poursat. A note on bic in mixed-effects models. *Electronic Journal of Statistics*, 8(1):456–475, January 2014. URL https://doi.org/10.1214/14-EJS890.

[46] H. J. Keselman, J. Algina, R. K. Kowalchuk, and R. D. Wolfinger. A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements. *Communications in Statistics-Simulation and computation*, 27(3):591–604, 1998.

[47] M. J. Gurka. Selecting the best linear mixed model under reml. *The American Statistician*, 60(1):19–26, 2006.

[48] J. G. Ibrahim, H. Zhu, and N. Tang. Model selection criteria for missing-data problems using the em algorithm. *Journal of the American Statistical Association*, 103(484): 1648–1658, December 2008. ISSN 1537-274X. URL https://pubmed.ncbi.nlm.nih.gov/19693282.

[49] M. J. Lindstrom and D. M. Bates. Newton-raphson and em algorithms for linear mixed-effects models for repeated-measures data. *Journal of the American Statistical Association*, 83(404):1014–1022, 1988. ISSN 01621459. URL http://www.jstor.org/stable/2290128.

[50] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, Articles*, 67(1):1–48, 2015. ISSN 1548-7660. doi: 10.18637/jss.v067.i01. URL https://www.jstatsoft.org/v067/i01.

[51] E. A. Nadaraya. On estimating regression. *Theory of Probability & Its Applications*, 9 (1):141–142, 1964. doi: 10.1137/1109020. URL https://doi.org/10.1137/1109020.

[52] G. S. Watson. Smooth regression analysis. *Sankhya: The Indian Journal of Statistics, Series A (1961-2002)*, 26(4):359–372, 1964. ISSN 0581572X. URL http://www.jstor.org/stable/25049340.

[53] D. R. Hoover, J. A. Rice, C. O. Wu, and L.-P. Yang. Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85 (4):809–822, 12 1998. ISSN 0006-3444. doi: 10.1093/biomet/85.4.809. URL https://doi.org/10.1093/biomet/85.4.809.

[54] P. Craven and G. Wahba. Smoothing noisy data with spline functions. *Numerische Mathematik*, 31(4):377–403, December 1978. ISSN 0945-3245. URL https://doi.org/10.1007/BF01404567.

[55] J. Z. Huang, C. O. Wu, and L. Zhou. Polynomial spline estimation and inference for varying coefficient models with longitudinal data. *Statistica Sinica*, 14(3):763–788, 2004. ISSN 10170405, 19968507. URL http://www.jstor.org/stable/24307415.

[56] J. Z. Huang, C. O. Wu, and L. Zhou. Varying-coefficient models and basis function approximations for the analysis of repeated measurements. *Biometrika*, 89(1):111–128, 2002. ISSN 00063444. URL http://www.jstor.org/stable/4140562.

[57] A. Verhasselt. Generalized varying coefficient models: A smooth variable selection technique. *Statistica Sinica*, 24(1):147–171, 2014. ISSN 10170405, 19968507. URL http://www.jstor.org/stable/26432537.

[58] Z. Cai, J. Fan, and R. Li. Efficient estimation and inferences for varying-coefficient models. *Journal of the American Statistical Association*, 95(451):888–902, 2000. doi: 10.1080/01621459.2000.10474280. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.2000.10474280.

[59] D. Şentürk, L. S. Dalrymple, S. M. Mohammed, G. A. Kaysen, and D. V. Nguyen. Modeling time-varying effects with generalized and unsynchronized longitudinal data. *Statistics in medicine*, 32(17):2971–2987, July 2013. ISSN 0277-6715. URL https://pubmed.ncbi.nlm.nih.gov/23335196.

[60] T. Hastie and R. Tibshirani. Generalized additive models for medical research. *Statistical Methods in Medical Research*, 4:187 – 196, 1995.

[61] T. Hastie and R. Tibshirani. Generalized Additive Models. *Statistical Science*, 1(3):297 – 310, 1986. doi: 10.1214/ss/1177013604. URL https://doi.org/10.1214/ss/1177013604.

[62] C. O. Wu, C.-T. Chiang, and D. R. Hoover. Asymptotic confidence regions for kernel smoothing of a varying-coefficient model with longitudinal data. *Journal of the American Statistical Association*, 93(444):1388–1402, 1998. doi: 10.1080/01621459.1998.10473800. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1998.10473800.

[63] J. D. Hart and T. E. Wehrly. Consistency of cross-validation when the data are curves. *Stochastic Processes and their Applications*, 45(2):351 – 361, 1993. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(93)90080-N. URL http://www.sciencedirect.com/science/article/pii/030441499390080N.

[64] C. O. Wu and C.-T. Chiang. Kernel smoothing on varying coefficient models with longitudinal dependent variable. *Statistica Sinica*, pages 433–456, 2000.

[65] C.-T. Chiang, J. A. Rice, and C. O. Wu. Smoothing spline estimation for varying coefficient models with repeatedly measured dependent variables. *Journal of the American Statistical Association*, 96(454):605–619, 2001.

[66] Y. Li, D. V. Nguyen, E. Kürüm, C. M. Rhee, Y. Chen, K. Kalantar-Zadeh, and D. Şentürk. A multilevel mixed effects varying coefficient model with multilevel predictors and random effects for modeling hospitalization risk in patients on dialysis. *Biometrics*, 76(3):924–938, September 2020. ISSN 0006-341X. URL https://pubmed.ncbi.nlm.nih.gov/31856300.

[67] D. Zhang. Generalized linear mixed models with varying coefficients for longitudinal data. *Biometrics*, 60(1):8–15, 2004. ISSN 0006341X, 15410420. URL http://www.jstor.org/stable/3695546.

[68] J. D. Nielsen and C. B. Dean. Clustered mixed nonhomogeneous poisson process spline models for the analysis of recurrent event panel data. *Biometrics*, 64(3):751–761, September 2008. ISSN 0006-341X. URL https://pubmed.ncbi.nlm.nih.gov/18047528.

[69] X. Cai, L. Xue, X. Pu, and X. Yan. Efficient estimation for varying-coefficient mixed effects models with functional response data. *Metrika*, 84(4):467–495, May 2021. ISSN 1435-926X. URL https://doi.org/10.1007/s00184-020-00776-0.

[70] H. Wu and H. Liang. Backfitting random varying-coefficient models with time-dependent smoothing covariates. *Scandinavian Journal of Statistics*, 31(1):3–19, 2004. ISSN 03036898, 14679469. doi: https://doi.org/10.1111/j.1467-9469.2004.00369.x. URL http://www.jstor.org/stable/4616808.

[71] J. Fan and J.-T. Zhang. Two-step estimation of functional linear models with applications to longitudinal data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 62(2):303–322, 2000. ISSN 13697412, 14679868. doi: https://doi.org/10.1111/1467-9868.00233. URL http://www.jstor.org/stable/3088861.

[72] A. Buja, T. Hastie, and R. Tibshirani. Linear smoothers and additive models. *The Annals of Statistics*, 17(2):453–510, 1989. ISSN 00905364. doi: 10.1214/aos/1176347115. URL http://www.jstor.org/stable/2241560.

[73] J. Fletcher. What is heterogeneity and is it important? *BMJ*, 334(7584):94–96, 2007. ISSN 0959-8138. doi: 10.1136/bmj.39057.406644.68. URL https://www.bmj.com/content/334/7584/94.

[74] R. E. Quandt. A new approach to estimating switching regressions. *Journal of the American Statistical Association*, 67(338):306–310, 1972. ISSN 01621459. URL http://www.jstor.org/stable/2284373.

[75] R. E. Quandt and J. B. Ramsey. Estimating mixtures of normal distributions and switching regressions. *Journal of the American Statistical Association*, 73(364):730–738, 1978. doi: 10.1080/01621459.1978.10480085. URL https://www.tandfonline.com/doi/abs/10.1080/01621459.1978.10480085.

[76] G. J. McLachlan and D. Peel. *Finite mixture models*. Wiley Series in Probability and Statistics, New York, 2000.

[77] Y. Zou, Y. Zhang, and D. Lord. Application of finite mixture of negative binomial regression models with varying weight parameters for vehicle crash data analysis. *Accident Analysis & Prevention*, 50:1042–1051, 2013. ISSN 0001-4575. doi: https://doi.org/10.1016/j.aap.2012.08.004. URL https://www.sciencedirect.com/science/article/pii/S0001457512002916.

[78] K. L. Montgomery, M. G. Vaughn, S. J. Thompson, and M. O. Howard. Heterogeneity in drug abuse among juvenile offenders: Is mixture regression more informative than standard regression? *International Journal of Offender Therapy and Comparative Criminology*, 57(11):1326–1346, 2013. doi: 10.1177/0306624X12459185. URL https://doi.org/10.1177/0306624X12459185. PMID: 23027831.

[79] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38, 1977. ISSN 00359246. URL http://www.jstor.org/stable/2984875.

[80] C. Proust and H. Jacqmin-Gadda. Estimation of linear mixed models with a mixture of distribution for the random effects. *Computer methods and programs in biomedicine*, 78(2):165–173, May 2005. ISSN 0169-2607. doi: 10.1016/j.cmpb.2004.12.004. URL https://pubmed.ncbi.nlm.nih.gov/15848271.

[81] G. Celeux, O. Martin, and C. Lavergne. Mixture of linear mixed models for clustering gene expression profiles from repeated microarray experiments. *Statistical Modelling*, 5(3):243–267, 2005. doi: 10.1191/1471082X05st096oa. URL https://doi.org/10.1191/1471082X05st096oa.

[82] M. L. Van Horn, Y. Feng, M. Kim, A. Lamont, D. Feaster, and T. Jaki. Using multilevel regression mixture models to identify level-1 heterogeneity in level-2 effects. *Structural equation modeling : a multidisciplinary journal*, 23(2):259–269, 2016. ISSN 1532-8007. URL https://pubmed.ncbi.nlm.nih.gov/27274654.

[83] B. Muthén and T. Asparouhov. Multilevel regression mixture analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 172(3):639–657, 2009. ISSN 1467-985X. doi: 10.1111/j.1467-985X.2009.00589.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-985X.2009.00589.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-985X.2009.00589.x.

[84] C. Proust-Lima, V. Philipps, and B. Liquet. Estimation of extended mixed models using latent classes and latent processes: The r package lcmm. *Journal of Statistical Software, Articles*, 78(2):1–56, 2017. ISSN 1548-7660. doi: 10.18637/jss.v078.i02. URL https://www.jstatsoft.org/v078/i02.

[85] M. Huang, W. Yao, S. Wang, and Y. Chen. Statistical inference and applications of mixture of varying coefficient models: Mixture of varying coefficient models. *Scandinavian Journal of Statistics*, 45(3), February 2018. ISSN 0303-6898. doi: 10.1111/sjos.12316.

[86] R. A. Johnson, D. W. Wichern, et al. *Applied multivariate statistical analysis*, volume 6. Pearson London, UK:, 2014.

[87] J. P. Stevens. *Applied multivariate statistics for the social sciences, 4th ed.* Lawrence Erlbaum Associates Publishers, Mahwah, NJ, US, 2002.

[88] F. B. Bryant and P. R. Yarnold. Principal-components analysis and exploratory and confirmatory factor analysis. pages 99–136. American Psychological Association, Washington, DC, US, 1995.

[89] L. Klem. Structural equation modeling. pages 227–260. American Psychological Association, Washington, DC, US, 2000.

[90] S. Wright. The method of path coefficients. *The Annals of Mathematical Statistics*, 5 (3):161–215, 1934. ISSN 00034851. URL http://www.jstor.org/stable/2957502.

[91] L. M. Wolfle. Sewall wright on the method of path coefficients: An annotated bibliography. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(3): 280–291, 1999. doi: 10.1080/10705519909540134. URL https://doi.org/10.1080/10705519909540134.

[92] D. N. Lawley. Vi.—the estimation of factor loadings by the method of maximum likelihood. *Proceedings of the Royal Society of Edinburgh*, 60(1):64–82, 1940. doi: 10.1017/S037016460002006X.

[93] D. B. Rubin and D. T. Thayer. Em algorithms for ml factor analysis. *Psychometrika*, 47(1):69–76, March 1982. ISSN 1860-0980. URL https://doi.org/10.1007/BF02293851.

[94] J.-H. Zhao, P. L. H. Yu, and Q. Jiang. Ml estimation for factor analysis: Em or non-em ? *Statistics and Computing*, 18(2):109–123, June 2008. ISSN 1573-1375. URL https://doi.org/10.1007/s11222-007-9042-y.

[95] T. Asparouhov, E. L. Hamaker, and B. Muthén. Dynamic structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3):359–388, 2018. doi: 10.1080/10705511.2017.1406803. URL https://doi.org/10.1080/10705511.2017.1406803.

[96] R. Van De Schoot, P. Schmidt, A. De Beuckelaer, K. Lek, and M. Zondervan-Zwijnenburg. Editorial: Measurement invariance. *Frontiers in Psychology*, 6:1064, 2015. ISSN 1664-1078. doi: 10.3389/fpsyg.2015.01064. URL https://www.frontiersin.org/article/10.3389/fpsyg.2015.01064.

[97] D. L. Putnick and M. H. Bornstein. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41:71–90, 2016. ISSN 0273-2297. doi: https://doi.org/10.1016/j.dr.2016.06.004. URL https://www.sciencedirect.com/science/article/pii/S0273229716300351.

[98] N. Christelis, B. Simpson, M. Russo, M. Stanton-Hicks, G. Barolat, S. Thomson, S. Schug, R. Baron, E. Buchser, D. B. Carr, T. R. Deer, I. Dones, S. Eldabe, R. Gallagher, F. Huygen, D. Kloth, R. Levy, R. North, C. Perruchoud, E. Petersen, P. Rigoard, K. Slavin, D. Turk, T. Wetzel, and J. Loeser. Persistent Spinal Pain Syndrome: a proposal for Failed Back Surgery Syndrome and ICD-11. *Pain Medicine*, 01 2021. ISSN 1526-4637. doi: 10.1093/pm/pnab015. URL https://doi.org/10.1093/pm/pnab015. pnab015.

[99] L. Scalone, F. Zucco, A. Lavano, A. Costantini, M. De Rose, P. Poli, G. Fortini, L. Demartini, E. De Simone, V. Menardo, M. Meglio, P. Cozzolino, P. A. Cortesi, and L. G. Mantovani. Benefits in pain perception, ability function and health-related quality of life in patients with failed back surgery syndrome undergoing spinal cord stimulation in a clinical practice setting. *Health and Quality of Life Outcomes*, 16(1): 68, April 2018. ISSN 1477-7525. URL https://doi.org/10.1186/s12955-018-0887-x.

[100] L. Castelli, V. Tesio, F. Colonna, S. Molinaro, P. Leombruni, M. Bruzzone, E. Fusaro, P. Sarzi-Puttini, and R. Torta. Alexithymia and psychological distress in fibromyalgia:

prevalence and relation with quality of life. *Clinical and Experimental Rheumatology*, 30(6 Suppl 74):70–77, December 2012. ISSN 0392-856X.

[101] M. Davies, S. Brophy, R. Williams, and A. Taylor. The prevalence, severity, and impact of painful diabetic peripheral neuropathy in type 2 diabetes. *Diabetes Care*, 29 (7):1518–1522, July 2006. ISSN 0149-5992. doi: 10.2337/dc05-2228.

[102] M. I. Hasenbring, H. Plaas, B. Fischbein, and R. Willburger. The relationship between activity and pain in patients 6 months after lumbar disc surgery: do pain-related coping modes act as moderator variables ? *European Journal of Pain (London, England)*, 10 (8):701–709, November 2006. ISSN 1090-3801. doi: 10.1016/j.ejpain.2005.11.004.

[103] P. Langley, C. Pérez Hernández, C. Margarit Ferri, D. Ruiz Hidalgo, and M. Lubián López. Pain, health related quality of life and healthcare resource utilization in Spain. *Journal of Medical Economics*, 14(5):628–638, 2011. ISSN 1941-837X. doi: 10.3111/13696998.2011.604369.

[104] F. M. Kovacs, V. Abraira, J. Zamora, M. Teresa Gil del Real, J. Llobera, C. Fernández, and the Kovacs-Atención Primaria Group. Correlation between pain, disability, and quality of life in patients with common low back pain. *Spine*, 29(2), 2004. ISSN 0362-2436. URL https://journals.lww.com/spinejournal/Fulltext/2004/01150/Correlation_Between_Pain,_Disability,_and_Quality.18.aspx.

[105] E. Rapti, D. Damigos, P. Apostolara, V. Roka, C. Tzavara, and C. Lionis. Patients with chronic pain: evaluating depression and their quality of life in a single center study in greece. *BMC Psychology*, 7(1):86, December 2019. ISSN 2050-7283. URL https://doi.org/10.1186/s40359-019-0366-0.

[106] Y. Chen, P. Campbell, V. Y. Strauss, N. E. Foster, K. P. Jordan, and K. M. Dunn. Trajectories and predictors of the long-term course of low back pain: cohort study with 5-year follow-up. *Pain*, 159(2):252–260, 2018. ISSN 1872-6623. doi: 10.1097/j.pain.0000000000001097.

[107] K. M. Dunn, K. Jordan, and P. R. Croft. Characterizing the course of low back pain: a latent class analysis. *American Journal of Epidemiology*, 163(8):754–761, April 2006. ISSN 0002-9262. doi: 10.1093/aje/kwj100.

[108] M. Herdman, C. Gudex, A. Lloyd, M. Janssen, P. Kind, D. Parkin, G. Bonsel, and X. Badia. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research: An International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 20(10):1727–1736, December 2011. ISSN 1573-2649. doi: 10.1007/s11136-011-9903-x.

[109] M. Obradovic, A. Lal, and H. Liedgens. Validity and responsiveness of EuroQol-5 dimension (EQ-5D) versus Short Form-6 dimension (SF-6D) questionnaire in chronic pain. *Health and Quality of Life Outcomes*, 11:110, July 2013. ISSN 1477-7525. doi: 10.1186/1477-7525-11-110.

[110] B. X. Tran, A. Ohinmaa, and L. T. Nguyen. Quality of life profile and psychometric properties of the EQ-5D-5L in HIV/AIDS patients. *Health and Quality of Life Outcomes*, 10:132, November 2012. ISSN 1477-7525. doi: 10.1186/1477-7525-10-132.

[111] P. Vartiainen, P. Mäntyselkä, T. Heiskanen, N. Hagelberg, S. Mustola, H. Forssell, H. Kautiainen, and E. Kalso. Validation of EQ-5D and 15D in the assessment of health-related quality of life in chronic pain. *Pain*, 158(8):1577–1585, 2017. ISSN 1872-6623. doi: 10.1097/j.pain.0000000000000954.

[112] M. A. Ferreira-Valente, J. L. Pais-Ribeiro, and M. P. Jensen. Validity of four pain intensity rating scales. *Pain*, 152(10):2399–2404, October 2011. ISSN 1872-6623. doi: 10.1016/j.pain.2011.07.005.

[113] J. C. Fairbank, J. Couper, J. B. Davies, and J. P. O'Brien. The Oswestry low back pain disability questionnaire. *Physiotherapy*, 66(8):271–273, August 1980. ISSN 0031-9406.

[114] C.-P. Lee, T.-S. Fu, C.-Y. Liu, and C.-I. Hung. Psychometric evaluation of the Oswestry Disability Index in patients with chronic low back pain: factor and Mokken analyses. *Health and Quality of Life Outcomes*, 15, October 2017. ISSN 1477-7525. doi: 10.1186/s12955-017-0768-8. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5627480/.

[115] M. Vianin. Psychometric properties and clinical usefulness of the Oswestry Disability Index. *Journal of Chiropractic Medicine*, 7(4):161–163, December 2008. ISSN 1556-3707. doi: 10.1016/j.jcm.2008.07.001. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2697602/.

[116] A. S. Zigmond and R. P. Snaith. The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica*, 67(6):361–370, June 1983. ISSN 0001-690X. doi: 10.1111/j.1600-0447.1983.tb09716.x.

[117] D. C. Turk, R. H. Dworkin, J. J. Trudeau, C. Benson, D. M. Biondi, N. P. Katz, and M. Kim. Validation of the Hospital Anxiety and Depression Scale in Patients With Acute Low Back Pain. *The Journal of Pain: Official Journal of the American Pain Society*, 16(10):1012–1021, October 2015. ISSN 1528-8447. doi: 10.1016/j.jpain.2015.07.001.

[118] I. Bjelland, A. A. Dahl, T. T. Haug, and D. Neckelmann. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *Journal of Psychosomatic Research*, 52(2):69–77, February 2002. ISSN 0022-3999. doi: 10.1016/s0022-3999(01)00296-3.

[119] M. A. Sagheer, M. F. Khan, and S. Sharif. Association between chronic low back pain, anxiety and depression in patients at a tertiary care centre. *JPMA. The Journal of the Pakistan Medical Association*, 63(6):688–690, June 2013. ISSN 0030-9982.

[120] S. Irachabal, M. Koleck, N. Rascle, and M. Bruchon-Schweitzer. [Pain coping strategies: French adaptation of the coping strategies questionnaire (CSQ-F)]. *L'Encephale*, 34(1):47–53, January 2008. ISSN 0013-7006. doi: 10.1016/j.encep.2006.11.002.

[121] B. Efron. Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics*, 7(1):1–26, 1979. ISSN 0090-5364. URL https://www.jstor.org/stable/2958830. Publisher: Institute of Mathematical Statistics.

[122] H. Shono. Efficiency of the finite correction of Akaike's Information Criteria. *Fisheries Science*, 66(3):608–610, 2000. ISSN 1444-2906. doi: https://doi.org/10.1046/j.1444-2906.2000.00095.x. URL https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1444-2906.2000.00095.x. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1046/j.1444-2906.2000.00095.x.

[123] R. J. Gatchel, Y. B. Peng, M. L. Peters, P. N. Fuchs, and D. C. Turk. The biopsychosocial approach to chronic pain: scientific advances and future directions. *Psychological Bulletin*, 133(4):581–624, July 2007. ISSN 0033-2909. doi: 10.1037/0033-2909.133.4.581.

[124] N. Naiditch, M. Billot, M. Moens, L. Goudman, P. Cornet, D. Le Breton, M. Roulaud, A. Ounajim, P. Page, B. Lorgeoux, K. Nivole, P. Pries, C. Swennen, S. Teyssedou, E. Charrier, G. B. de Montgazon, P. F. Descoins, B. Roy-Moreau, N. Grimaud, R. David, T. Vendeuvre, and P. Rigoard. Persistent Spinal Pain Syndrome Type 2 (PSPS-T2), a Social Pain ? Advocacy for a Social Gradient of Health Approach to Chronic Pain. *Journal of Clinical Medicine*, 10(13):2817, January 2021. doi: 10.3390/jcm10132817. URL https://www.mdpi.com/2077-0383/10/13/2817. Number: 13 Publisher: Multidisciplinary Digital Publishing Institute.

[125] L. Goudman and M. Moens. Moving Beyond a Pain Intensity Reporting: The Value of Goal Identification in Neuromodulation. *Neuromodulation: Technology at the Neural Interface*, 23(7):1057–1058, 2020. ISSN 1525-1403. doi: https://doi.org/10.1111/ner.13151. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/ner.13151. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/ner.13151.

[126] S. J. Kamper, A. T. Apeldoorn, A. Chiarotto, R. J. E. M. Smeets, R. W. J. G. Ostelo, J. Guzman, and M. W. van Tulder. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. *The Cochrane Database of Systematic Reviews*, (9):CD000963, September 2014. ISSN 1469-493X. doi: 10.1002/14651858.CD000963.pub3.

[127] D. Marris, K. Theophanous, P. Cabezon, Z. Dunlap, and M. Donaldson. The impact of combining pain education strategies with physical therapy interventions for patients with chronic pain: A systematic review and meta-analysis of randomized controlled trials. *Physiotherapy Theory and Practice*, 37(4):461–472, April 2021. ISSN 0959-3985. doi: 10.1080/09593985.2019.1633714. URL https://doi.org/10.1080/09593985.2019.1633714. Publisher: Taylor & Francis _eprint: https://doi.org/10.1080/09593985.2019.1633714.

[128] R. S. Roth and M. E. Geisser. Educational achievement and chronic pain disability: mediating role of pain-related cognitions. *The Clinical Journal of Pain*, 18(5):286–296, October 2002. ISSN 0749-8047. doi: 10.1097/00002508-200209000-00003.

[129] A. Leclerc, J. Gourmelen, J.-F. Chastang, S. Plouvier, I. Niedhammer, and J.-L. Lanoë. Level of education and back pain in France: the role of demographic, lifestyle and physical work factors. *International Archives of Occupational and Environmental Health*, 82(5):643–652, April 2009. ISSN 1432-1246. doi: 10.1007/s00420-008-0375-4.

[130] A. Samulowitz, I. Gremyr, E. Eriksson, and G. Hensing. "Brave Men" and "Emotional Women": A Theory-Guided Literature Review on Gender Bias in Health Care and Gendered Norms towards Patients with Chronic Pain. *Pain Research & Management*, 2018:6358624, 2018. ISSN 1918-1523. doi: 10.1155/2018/6358624.

[131] D. Higuchi. Adaptive and maladaptive coping strategies in older adults with chronic pain after lumbar surgery. *International Journal of Rehabilitation Research*, 43 (2):116–122, June 2020. ISSN 0342-5282. doi: 10.1097/MRR.0000000000000389. URL https://journals.lww.com/intjrehabilres/Abstract/2020/06000/Adaptive_and_maladaptive_coping_strategies_in.3.aspx.

[132] S. G. Henry, R. A. Bell, J. J. Fenton, and R. L. Kravitz. Goals of Chronic Pain Management: Do Patients and Primary Care Physicians Agree and Does it Matter? *The Clinical Journal of Pain*, 33(11):955–961, November 2017. ISSN 1536-5409. doi: 10.1097/AJP.0000000000000488.

[133] K. R. Archer, C. J. Devin, S. W. Vanston, T. Koyama, S. E. Phillips, S. L. Mathis, S. Z. George, M. J. McGirt, D. M. Spengler, O. S. Aaronson, J. S. Cheng, and S. T. Wegener. Cognitive-Behavioral-Based Physical Therapy for Patients With Chronic Pain Undergoing Lumbar Spine Surgery: A Randomized Controlled Trial. *The Journal of Pain*, 17(1):76–89, 2016. ISSN 1528-8447. doi: 10.1016/j.jpain.2015.09.013.

[134] K. J. Barnhoorn, H. van de Meent, R. T. M. van Dongen, F. P. Klomp, H. Groenewoud, H. Samwel, M. W. G. Nijhuis-van der Sanden, J. P. M. Frölke, and J. B. Staal. Pain exposure physical therapy (PEPT) compared to conventional treatment in complex regional pain syndrome type 1: a randomised controlled trial. *BMJ open*, 5(12):e008283, December 2015. ISSN 2044-6055. doi: 10.1136/bmjopen-2015-008283.

[135] X. Lin and R. J. Carroll. Semiparametric regression for clustered data. *Biometrika*, 88 (4):1179–1185, 2001. ISSN 00063444. doi: https://doi.org/10.1093/biomet/88.4.1179. URL http://www.jstor.org/stable/2673710.

[136] Y. Sun, W. Zhang, and H. Tong. Estimation of the covariance matrix of random effects in longitudinal studies. *Ann. Statist.*, 35(6):2795–2814, 12 2007. doi: 10.1214/009053607000000523. URL https://doi.org/10.1214/009053607000000523.

[137] L. Su and M. J. Daniels. Bayesian modeling of the covariance structure for irregular longitudinal data using the partial autocorrelation function. *Statistics in medicine*, 34(12):2004–2018, May 2015. ISSN 0277-6715. doi: 10.1002/sim.6465. URL https://pubmed.ncbi.nlm.nih.gov/25762065.

[138] M. J. Gurka, L. J. Edwards, and K. E. Muller. Avoiding bias in mixed model inference for fixed effects. *Statistics in medicine*, 30(22):2696–2707, September 2011. ISSN 0277-6715. URL https://pubmed.ncbi.nlm.nih.gov/21751227.

[139] A. Gelman and J. Hill. Data analysis using regression and hierarchical/multilevel models. *New York, NY: Cambridge*, 2007.

[140] M. Stephens. Dealing with label switching in mixture models. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 62(4):795–809, 2000. ISSN 13697412, 14679868. doi: https://doi.org/10.1111/1467-9868.00265. URL http://www.jstor.org/stable/2680622.

[141] L. Breiman and J. H. Friedman. Estimating optimal transformations for multiple regression and correlation. *Journal of the American Statistical Association*, 80(391): 580–598, 1985. ISSN 01621459. doi: 10.2307/2288477. URL http://www.jstor.org/stable/2288473.

[142] E. Shireman, D. Steinley, and M. J. Brusco. Examining the effect of initialization strategies on the performance of gaussian mixture modeling. *Behavior Research Methods*, 49(1):282–293, February 2017. ISSN 1554-3528. doi: https://doi.org/10.3758/s13428-015-0697-6.

[143] N. Laird. Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73(364):805–811, 1978. doi: https://doi.org/10.1080/01621459.1978.10480103.

[144] D. Böhning, E. Dietz, R. Schaub, P. Schlattmann, and B. G. Lindsay. The distribution of the likelihood ratio for mixtures of densities from the one-parameter exponential family. *Annals of the Institute of Statistical Mathematics*, 46(2):373–388, 1994. doi: https://doi.org/10.1007/BF01720593.

[145] Y. Yang. Can the strengths of aic and bic be shared? a conflict between model indentification and regression estimation. *Biometrika*, 92(4):937–950, 2005. ISSN 00063444. doi: 10.1093/biomet/92.4.937. URL http://www.jstor.org/stable/20441246.

[146] P. Orbanz and Y. W. Teh. *Bayesian Nonparametric Models*, pages 81–89. Springer US, Boston, MA, 2010. ISBN 978-0-387-30164-8. doi: 10.1007/978-0-387-30164-8_66. URL https://doi.org/10.1007/978-0-387-30164-8_66.

[147] J.-M. Marin, K. L. Mengersen, and C. P. Robert. Bayesian modelling and inference on mixtures of distributions. *Handbook of Statistics*, 25:459–507, 2005. doi: https://doi.org/10.1016/S0169-7161(05)25016-2.

[148] J. Fan, C. Zhang, and J. Zhang. Generalized likelihood ratio statistics and wilks phenomenon. *Ann. Statist.*, 29(1):153–193, 02 2001. doi: 10.1214/aos/996986505. URL https://doi.org/10.1214/aos/996986505.

[149] L. Xue, X. Shu, P. Shi, C. Wu, and A. Qu. Time-varying feature selection for longitudinal analysis. *Statistics in medicine*, 2019. doi: https://doi.org/10.1002/sim.8412.

[150] Q. Li and J. S. Racine. *Nonparametric Econometrics: Theory and Practice*. Number 8355 in Economics Books. Princeton University Press, 2006. doi: https://doi.org/10.1111/j.1475-4932.2011.00750.x. URL https://ideas.repec.org/b/pup/pbooks/8355.html.

[151] W. Zhang and S.-Y. Lee. Variable Bandwidth Selection in Varying-Coefficient Models. *Journal of Multivariate Analysis*, 74(1):116–134, July 2000. doi: https://doi.org/10.1006/jmva.1999.1883. URL https://ideas.repec.org/a/eee/jmvana/v74y2000i1p116-134.html.

[152] Y. Slaoui and A. Jmaei. Recursive density estimators based on robbins-monro's scheme and using bernstein polynomials. *Statistics and Its Interface*, 12(3):439–455, 2019. doi: 10.4310/19-sii561.

[153] S. Helali and Y. Slaoui. Estimation of a distribution function using lagrange polynomials with tchebychev-gauss points. *Statistics and Its Interface*, 13(3):399–410, 2020. doi: 10.4310/SII.2020.v13.n3.a9.

[154] P. Hall and B. U. Park. New methods for bias correction at endpoints and boundaries. *The Annals of Statistics*, 30(5):1460–1479, 2002. doi: 10.1214/aos/1035844983.

[155] I. B. Wilson and P. D. Cleary. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA*, 273(1):59–65, January 1995. ISSN 0098-7484.

[156] H. Breivik, E. Eisenberg, and T. O'Brien. The individual and societal burden of chronic pain in europe: the case for strategic prioritisation and action to improve knowledge and availability of appropriate care. *BMC Public Health*, 13(1):1229, December 2013. ISSN 1471-2458. URL https://doi.org/10.1186/1471-2458-13-1229.

[157] M. Wettstein, W. Eich, C. Bieber, and J. Tesarz. Pain Intensity, Disability, and Quality of Life in Patients with Chronic Low Back Pain: Does Age Matter ? *Pain Medicine (Malden, Mass.)*, 20(3):464–475, 2019. ISSN 1526-4637. doi: 10.1093/pm/pny062.

[158] J. C. Pinheiro and D. M. Bates. *Mixed-effects models in S and S-PLUS*. Springer, New York, NY [u.a.], 2000. ISBN 0387989579 9780387989570 9781441903174 1441903178. doi: 10.1007/978-1-4419-0318-1. URL http://www.worldcat.org/search?qt=worldcat_org_all&q=1441903178.

[159] C. M. Crainiceanu and D. Ruppert. Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, 66(1):165–185, 2004. ISSN 13697412, 14679868. doi: https://doi.org/10.1111/j.1467-9868.2004.00438.x. URL http://www.jstor.org/stable/3647633.

[160] S. Greven, C. M. Crainiceanu, H. Küchenhoff, and A. Peters. Restricted likelihood ratio testing for zero variance components in linear mixed models. *Journal of Computational and Graphical Statistics*, 17(4):870–891, 2008. doi: https://doi.org/10.1198/106186008X386599.

[161] F. Cosci and G. A. Fava. Staging of mental disorders: Systematic review, 2013. ISSN 0033-3190. URL https://www.karger.com/DOI/10.1159/000342243.

[162] C. M. Butowicz, S. P. Silfies, J. Vendemia, S. Farrokhi, and B. D. Hendershot. Characterizing and Understanding the Low Back Pain Experience Among Persons with Lower Limb Loss. *Pain Medicine*, 21(5):1068–1077, 11 2019. ISSN 1526-2375. doi: 10.1093/pm/pnz293. URL https://doi.org/10.1093/pm/pnz293.

[163] K. D. Roover. Finding clusters of groups with measurement invariance: Unraveling intercept non-invariance with mixture multigroup factor analysis. *Structural Equation Modeling: A Multidisciplinary Journal*, 28(5):663–683, 2021. doi: 10.1080/10705511.2020.1866577. URL https://doi.org/10.1080/10705511.2020.1866577.

[164] J. Baek and G. J. McLachlan. Mixtures of factor analyzers with common factor loadings for the clustering and visualisation of high-dimensional data. 2008.

[165] G. H. Lubke and B. Muthén. Investigating population heterogeneity with factor mixture models. *Psychological methods*, 10 1:21–39, 2005.

[166] K. E. Masyn, H. Petras, and W. Liu. *Growth Curve Models with Categorical Outcomes*, pages 2013–2025. Springer New York, New York, NY, 2014. ISBN 978-1-4614-5690-2. doi: 10.1007/978-1-4614-5690-2_404. URL https://doi.org/10.1007/978-1-4614-5690-2_404.

[167] S. Wasserman, A. von Eye, and C. C. Clogg. Categorical variables in developmental research : methods of analysis. *Journal of the American Statistical Association*, 94: 1384, 1999.

[168] C. Liu and D. B. Rubin. Maximum likelihood estimation of factor analysis using the ecme algorithm with complete and incomplete data. 1998.

[169] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis. A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems. *IEEE Transactions on Signal Processing*, 64:76–88, 2016.

[170] C. Liu, D. B. Rubin, and Y. N. Wu. Parameter expansion to accelerate em: The px-em algorithm. *Biometrika*, 85:755–770, 1998.

[171] M. Forni, M. Hallin, M. Lippi, and L. Reichlin. The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, 82:540–554, 2000.

[172] L. Su and X. Wang. On time-varying factor models: Estimation and testing. *Journal of Econometrics*, 198(1):84–101, 2017. ISSN 0304-4076. doi: https://doi.org/10.1016/j.jeconom.2016.12.004. URL https://www.sciencedirect.com/science/article/pii/S0304407617300143.

[173] J. Lethem, P. Slade, J. Troup, and G. Bentley. Outline of a fear-avoidance model of exaggerated pain perception—i. *Behaviour Research and Therapy*, 21(4):401–408, 1983. ISSN 0005-7967. doi: https://doi.org/10.1016/0005-7967(83)90009-8. URL https://www.sciencedirect.com/science/article/pii/0005796783900098. Special Announcement.

[174] F. M. Kovacs, A. Muriel, V. Abriaira, J. M. Medina, M. D. Castillo Sanchez, J. Olabe, and Spanish Back Pain Research Network. The influence of fear avoidance beliefs on disability and quality of life is sparse in Spanish low back pain patients. *Spine*, 30(22): E676–682, November 2005. ISSN 1528-1159. doi: 10.1097/01.brs.0000186468.29359.e4.

[175] R. J. Vandenberg and C. E. Lance. A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3(1):4–70, 2000. doi: 10.1177/109442810031002. URL https://doi.org/10.1177/109442810031002.

[176] W. Meredith. Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4):525–543, December 1993. ISSN 1860-0980. URL https://doi.org/10.1007/BF02294825.

[177] T. L. Milfont and R. Fischer. Testing measurement invariance across groups: applications in cross-cultural research. *International journal of psychological research*, 3:111–130, 2010.

[178] G. Celeux, S. Fruewirth-Schnatter, and C. P. Robert. Model selection for mixture models - perspectives and strategies. *arXiv: Methodology*, 2018.

[179] J. Zhao, L. Jin, and L. Shi. Mixture model selection via hierarchical bic. *Computational Statistics & Data Analysis*, 88:139–153, 2015. ISSN 0167-9473. doi: https://doi.org/10.1016/j.csda.2015.01.019. URL https://www.sciencedirect.com/science/article/pii/S016794731500047X.

[180] R. Estabrook and M. Neale. A comparison of factor score estimation methods in the presence of missing data: Reliability and an application to nicotine dependence. *Multivariate behavioral research*, 48(1):1–27, January 2013. ISSN 0027-3171. URL https://pubmed.ncbi.nlm.nih.gov/24049215.

[181] P. J. Ferrando. Seven decades of factor analysis: From yela to the present day. *Psicothema*, 33:378+, July 2021. ISSN 02149915. URL https://link.gale.com/apps/doc/A671195660/HRCA?u=anon~dca9fdb&sid=googleScholar&xid=b4c8a4fd.

[182] K. V. Patel, D. Amtmann, M. P. Jensen, S. M. Smith, C. Veasley, and D. C. Turk. Clinical outcome assessment in clinical trials of chronic pain treatments. *Pain reports*, 6(1):e784–e784, January 2021. ISSN 2471-2531. URL https://pubmed.ncbi.nlm.nih.gov/33521482.

[183] K. V. Patel, R. Allen, L. Burke, J. T. Farrar, J. S. Gewandter, I. Gilron, N. P. Katz, J. D. Markman, S. F. Marshall, M. Resnick, A. S. C. Rice, M. C. Rowbotham, S. M.

Smith, G. F. Vanhove, A. D. Wasan, S. Zhang, R. H. Dworkin, and D. C. Turk. Evaluation of composite responder outcomes of pain intensity and physical function in neuropathic pain clinical trials: an acttion individual patient data analysis. *Pain*, 159 (11):2245–2254, November 2018. ISSN 0304-3959. URL https://pubmed.ncbi.nlm.nih.gov/30001225.

[184] P. Rigoard, M. Billot, P. Ingrand, I. Durand-Zaleski, M. Roulaud, P. Peruzzi, P. Dam Hieu, J. Voirin, S. Raoul, P. Page, M.-C. Djian, D. Fontaine, M. Lantéri-Minet, S. Blond, N. Buisset, E. Cuny, M. Cadenne, F. Caire, D. Ranoux, P. Mertens, H. Naous, E. Simon, E. Emery, G. Béraud, F. Debiais, G. Durand, A. Serrie, B. Diallo, J. Bulsei, A. Ounajim, K. Nivole, S. Duranton, N. Naiditch, O. Monlezun, and B. Bataille. How Should we Use Multicolumn Spinal Cord Stimulation to Optimize Back Pain Spatial Neural Targeting ? A Prospective, Multicenter, Randomized, Double-Blind, Controlled Trial (ESTIMET Study). *Neuromodulation: Journal of the International Neuromodulation Society*, August 2020. ISSN 1525-1403. doi: 10.1111/ner.13251.

[185] C. for Medicare & Medicaid Services. National coverage determination (NCD) for electrical nerve Stimulators. page 160:7, July 2021.

[186] L. Goudman, A. De Smedt, S. Eldabe, P. Rigoard, B. Linderoth, M. De Jaeger, M. Moens, and D. Consortium. High-dose spinal cord stimulation for patients with failed back surgery syndrome: a multicenter effectiveness and prediction study. *PAIN*, Articles in Press, December 2020. ISSN 0304-3959. doi: 10.1097/j.pain.0000000000002035. URL https://journals.lww.com/pain/Abstract/9000/High_dose_spinal_cord_stimulation_for_patients.98269.aspx.

[187] A. Ounajim, M. Billot, L. Goudman, P.-Y. Louis, Y. Slaoui, M. Roulaud, B. Bouche, P. Page, B. Lorgeoux, S. Baron, N. Adjali, K. Nivole, N. Naiditch, C. Wood, R. Rigoard, R. David, M. Moens, and P. Rigoard. Machine learning algorithms provide greater prediction of response to scs than lead screening trial: A predictive ai-based multicenter study. *Journal of Clinical Medicine*, 10(20), 2021. ISSN 2077-0383. doi: 10.3390/jcm10204764. URL https://www.mdpi.com/2077-0383/10/20/4764.

[188] P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 04 1983. ISSN 0006-3444. doi: 10.1093/biomet/70.1.41. URL https://doi.org/10.1093/biomet/70.1.41.

[189] R. Lenain, J. Boucquemont, K. Leffondré, C. Couchoud, M. Lassalle, M. Hazzan, and Y. Foucher. Clinical trial emulation by matching time-dependent propensity scores: The example of estimating impact of kidney transplantation. *Epidemiology*, 32(2), 2021. ISSN 1044-3983. URL https://journals.lww.com/epidem/Fulltext/2021/03000/Clinical_Trial_Emulation_by_Matching.9.aspx.

[190] P. Ganty and M. Sharma. Failed back surgery syndrome: a suggested algorithm of care. *British Journal of Pain*, 6(4):153–161, November 2012. ISSN 2049-4637. doi: 10.1177/2049463712470222. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4590103/.

[191] J. C. Leveque, A. T. Villavicencio, K. R. Bulsara, L. Rubin, and J. P. Gorecki. Spinal cord stimulation for failed back surgery syndrome. *Neuromodulation: Journal of the International Neuromodulation Society*, 4(1):1–9, January 2001. ISSN 1094-7159. doi: 10.1046/j.1525-1403.2001.00001.x.

[192] R. B. North, J. N. Campbell, C. S. James, M. K. Conover-Walker, H. Wang, S. Piantadosi, J. D. Rybock, and D. M. Long. Failed back surgery syndrome: 5-year follow-up in 102 patients undergoing repeated operation. *Neurosurgery*, 28(5):685–690 ; discussion 690–691, May 1991. ISSN 0148-396X.

[193] J. D. Law, R. A. Lehman, and W. M. Kirsch. Reoperation after lumbar intervertebral disc surgery. *Journal of Neurosurgery*, 48(2):259–263, February 1978. ISSN 0022-3085. doi: 10.3171/jns.1978.48.2.0259.

[194] T. R. Lehmann and H. S. LaRocca. Repeat lumbar surgery. A review of patients with failure from previous lumbar surgery treated by spinal canal exploration and lumbar spinal fusion. *Spine*, 6(6):615–619, December 1981. ISSN 0362-2436.

[195] P. J. Wiffen, S. Derry, R. F. Bell, A. S. Rice, T. R. Tölle, T. Phillips, and R. A. Moore. Gabapentin for chronic neuropathic pain in adults. *The Cochrane Database of Systematic Reviews*, 6:CD007938, 2017. ISSN 1469-493X. doi: 10.1002/14651858.CD007938.pub4.

[196] K. M. Dunn, K. W. Saunders, C. M. Rutter, C. J. Banta-Green, J. O. Merrill, M. D. Sullivan, C. M. Weisner, M. J. Silverberg, C. I. Campbell, B. M. Psaty, and M. Von Korff. Opioid prescriptions for chronic pain and overdose: a cohort study. *Annals of Internal Medicine*, 152(2):85–92, January 2010. ISSN 1539-3704. doi: 10.7326/0003-4819-152-2-201001190-00006.

[197] Y. Molero, H. Larsson, B. M. D'Onofrio, D. J. Sharp, and S. Fazel. Associations between gabapentinoids and suicidal behaviour, unintentional overdoses, injuries, road traffic incidents, and violent crime: population based cohort study in Sweden. *BMJ (Clinical research ed.)*, 365:l2147, 2019. ISSN 1756-1833. doi: 10.1136/bmj.l2147.

[198] S. H. Farber, J. L. Han, A. A. Elsamadicy, Q. Hussaini, S. Yang, P. Pagadala, B. Parente, J. Xie, and S. P. Lad. Long-term Cost Utility of Spinal Cord Stimulation in Patients with Failed Back Surgery Syndrome. *Pain Physician*, 20(6):E797–E805, 2017. ISSN 2150-1149.

[199] K. Kumar, R. S. Taylor, L. Jacques, S. Eldabe, M. Meglio, J. Molet, S. Thomson, J. O'Callaghan, E. Eisenberg, G. Milbouw, E. Buchser, G. Fortini, J. Richardson, and R. B. North. The effects of spinal cord stimulation in neuropathic pain are sustained: a 24-month follow-up of the prospective randomized controlled multicenter trial of the effectiveness of spinal cord stimulation. *Neurosurgery*, 63(4):762–770 ; discussion 770, October 2008. ISSN 1524-4040. doi: 10.1227/01.NEU.0000325731.46702.D9.

[200] C. M. Campbell, R. N. Jamison, and R. R. Edwards. Psychological Screening/Phenotyping as Predictors for Spinal Cord Stimulation. *Current pain*

and headache reports, 17(1):307, January 2013. ISSN 1531-3433. doi: 10.1007/s11916-012-0307-6. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3601592/.

[201] D. Bendersky and C. Yampolsky. Is spinal cord stimulation safe ? A review of its complications. *World Neurosurgery*, 82(6):1359–1368, December 2014. ISSN 1878-8769. doi: 10.1016/j.wneu.2013.06.012.

[202] E. Sparkes, J. H. Raphael, R. V. Duarte, K. LeMarchand, C. Jackson, and R. L. Ashford. A systematic literature review of psychological characteristics as determinants of outcome for spinal cord stimulation therapy. *Pain*, 150(2):284–289, August 2010. ISSN 1872-6623. doi: 10.1016/j.pain.2010.05.001.

[203] T. Wolter. Spinal cord stimulation for neuropathic pain: current perspectives. *Journal of Pain Research*, 7:651–663, 2014. ISSN 1178-7090. doi: 10.2147/JPR.S37589.

[204] T. Wolter, I. Fauler, and K. Kieselbach. The impact of psychological factors on outcomes for spinal cord stimulation: an analysis with long-term follow-up. *Pain Physician*, 16(3):265–275, June 2013. ISSN 2150-1149.

[205] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning. Springer Series in Statistics*. Springer New York Inc., New York, NY, USA, 2001.

[206] R. Brooks. EuroQol: the current state of play. *Health Policy (Amsterdam, Netherlands)*, 37(1):53–72, July 1996. ISSN 0168-8510. doi: 10.1016/0168-8510(96)00822-6.

[207] S. A. Montgomery and M. Asberg. A new depression scale designed to be sensitive to change. *The British Journal of Psychiatry: The Journal of Mental Science*, 134: 382–389, April 1979. ISSN 0007-1250. doi: 10.1192/bjp.134.4.382.

[208] J. Celestin, R. R. Edwards, and R. N. Jamison. Pretreatment psychosocial variables as predictors of outcomes following lumbar surgery and spinal cord stimulation: a systematic review and literature synthesis. *Pain Medicine (Malden, Mass.)*, 10(4): 639–653, June 2009. ISSN 1526-4637. doi: 10.1111/j.1526-4637.2009.00632.x.

[209] I. Strauss, K. Taha, V. Krishna, and M. Hodaie. Younger age predicts greater effectiveness of spinal cord stimulation for chronic pain. *Acta Neurochirurgica*, 158 (5):999–1003, May 2016. ISSN 0942-0940. doi: 10.1007/s00701-016-2753-0.

[210] J. Caylor, R. Reddy, S. Yin, C. Cui, M. Huang, C. Huang, R. Rao, D. G. Baker, A. Simmons, D. Souza, S. Narouze, R. Vallejo, and I. Lerman. Spinal cord stimulation in chronic pain: evidence and theory for mechanisms of action. *Bioelectronic Medicine*, 5(1):12, June 2019. ISSN 2332-8886. doi: 10.1186/s42234-019-0023-1. URL https://doi.org/10.1186/s42234-019-0023-1.

[211] T. Bendinger, N. Plunkett, D. Poole, and D. Turnbull. Psychological Factors as Outcome Predictors for Spinal Cord Stimulation. *Neuromodulation: Journal of the International Neuromodulation Society*, 18(6):465–471 ; discussion 471, August 2015. ISSN 1525-1403. doi: 10.1111/ner.12321.

[212] O. Marola, R. Cherala, J. Prusik, V. Kumar, C. Fama, M. Wilock, J. Crimmins, and J. G. Pilitsis. BMI as a Predictor of Spinal Cord Stimulation Success in Chronic Pain Patients. *Neuromodulation: Journal of the International Neuromodulation Society*, 20 (3):269–273, April 2017. ISSN 1525-1403. doi: 10.1111/ner.12482.

[213] F. van Eijs, H. Smits, J. W. Geurts, A. G. H. Kessels, M. A. Kemler, M. van Kleef, E. A. J. Joosten, and C. G. Faber. Brush-evoked allodynia predicts outcome of spinal cord stimulation in complex regional pain syndrome type 1. *European Journal of Pain (London, England)*, 14(2):164–169, February 2010. ISSN 1532-2149. doi: 10.1016/j.ejpain.2009.10.009.

[214] L. Mathew, C. Winfree, D. Miller-Saultz, and N. Sonty. Transcutaneous electrical nerve stimulator trial may be used as a screening tool prior to spinal cord stimulator implantation. *Pain*, 150(2):327–331, August 2010. ISSN 1872-6623. doi: 10.1016/j.pain.2010.05.018.

[215] K. Kumar, S. Rizvi, R. Nguyen, M. Abbas, S. Bishop, and V. Murthy. Impact of wait times on spinal cord stimulation therapy outcomes. *Pain Practice: The Official Journal of World Institute of Pain*, 14(8):709–720, November 2014. ISSN 1533-2500. doi: 10.1111/papr.12126.

[216] R. S. Taylor, M. J. Desai, P. Rigoard, and R. J. Taylor. Predictors of pain relief following spinal cord stimulation in chronic back and leg pain and failed back surgery syndrome: a systematic review and meta-regression analysis. *Pain Practice: The Official Journal of World Institute of Pain*, 14(6):489–505, July 2014. ISSN 1533-2500. doi: 10.1111/papr.12095.

[217] J. H. Friedman, T. Hastie, and R. Tibshirani. Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1):1–22, February 2010. ISSN 1548-7660. doi: 10.18637/jss.v033.i01. URL https://www.jstatsoft.org/index.php/jss/article/view/v033i01. Number: 1.

[218] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf. Support vector machines. *IEEE Intelligent Systems and their Applications*, 13(4):18–28, July 1998. ISSN 2374-9423. doi: 10.1109/5254.708428. Conference Name: IEEE Intelligent Systems and their Applications.

[219] D. Meyer, E. Dimitriadou, K. Hornik, A. Weingessel, F. Leisch, C.-C. C. l. C++-code), and C.-C. L. l. C++-code). e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien, October 2020. URL https://CRAN.R-project.org/package=e1071.

[220] N. Friedman, D. Geiger, and M. Goldszmidt. Bayesian Network Classifiers. *Machine Learning*, 29(2):131–163, November 1997. ISSN 1573-0565. doi: 10.1023/A:1007465528199. URL https://doi.org/10.1023/A:1007465528199.

[221] S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, USA, 2nd edition, 1998. ISBN 978-0-13-273350-2.

[222] T. B. Arnold. kerasR: R Interface to the Keras Deep Learning Library. *J. Open Source Softw.*, 2017. doi: 10.21105/joss.00296.

[223] T. Therneau, B. Atkinson, B. R. p. o. t. i. R. port, and maintainer 1999-2017). rpart: Recursive Partitioning and Regression Trees, April 2019. URL https://CRAN.R-project.org/package=rpart.

[224] L. Breiman. Random Forests. *Machine Learning*, 45(1):5–32, October 2001. ISSN 1573-0565. doi: 10.1023/A:1010933404324. URL https://doi.org/10.1023/A:1010933404324.

[225] M. N. Wright, S. Wager, and P. Probst. ranger: A Fast Implementation of Random Forests, January 2020. URL https://CRAN.R-project.org/package=ranger.

[226] J. H. Friedman. Stochastic gradient boosting. *Computational Statistics & Data Analysis*, 38(4):367–378, February 2002. ISSN 0167-9473. doi: 10.1016/S0167-9473(01)00065-2. URL https://doi.org/10.1016/S0167-9473(01)00065-2.

[227] T. Chen and C. Guestrin. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 785–794, August 2016. doi: 10.1145/2939672.2939785. URL http://arxiv.org/abs/1603.02754. arXiv: 1603.02754.

[228] E. Sparkes, R. V. Duarte, S. Mann, T. R. Lawrence, and J. H. Raphael. Analysis of psychological characteristics impacting spinal cord stimulation treatment outcomes: a prospective assessment. *Pain Physician*, 18(3):E369–377, June 2015. ISSN 2150-1149.

Mélange de coefficients aléatoires variant dans le temps et modèles d'analyse factorielle longitudinale et leur application à l'évaluation multidimensionnelle des douleurs chroniques

Résumé

Les données longitudinales multivariées jouent un rôle essentiel dans divers domaines de recherche, notamment les sciences médicales et comportementales. Elles permettent aux chercheurs de tester de multiples hypothèses, telles que l'identification de la variation dans le temps d'une variable clinique. La robustesse des inférences obtenues à partir de l'analyse des données longitudinales conduit à une complexité croissante des méthodes de traitement statistique. Cette complexité découle de la structure des données longitudinales et de leurs propriétés statistiques, notamment la dépendance temporelle intra-individuelle. Un autre degré de complexité peut provenir de l'hétérogénéité potentielle due à l'existence de plusieurs sous-populations latentes dans les données de l'étude. Actuellement, il est reconnu dans la littérature que la douleur présente un aspect longitudinal complexe. Il a été démontré que les corrélations entre les différentes dimensions de la douleur se renforcent avec le temps, non seulement lors du passage des douleurs aiguës aux douleurs chroniques, mais aussi au cours de l'évolution des douleurs chroniques. Cela peut être dû à l'accumulation à long terme des fardeaux biopsychosociales générées par la douleur chronique et aux altérations des mécanismes cérébraux reliant la douleur aux émotions. De plus, le caractère socioculturel et émotionnel de la douleur chronique rend son évaluation complexe en raison de l'hétérogénéité par rapport à l'importance des composantes sensorielles, émotionnelles et fonctionnelles dans la vie de chaque individu. Le premier objectif de cette thèse est de démontrer empiriquement l'existence de différents groupes latents de patients dont la qualité de vie est impactée différemment par les composantes sensorielles, émotionnelles et fonctionnelles de la douleur, par l'application d'un mélange de modèles à effet mixte sur des données longitudinales de patients souffrant de douleurs chroniques après une chirurgie rachidienne. Cela permet de justifier la nécessité d'une évaluation hétérogène et spécifique au patient. Le deuxième objectif est de proposer un nouveau modèle non paramétrique formulé comme un mélange de modèles à coefficients variables dans le temps incluant des processus aléatoires. Le modèle proposé a été ajusté avec un algorithme EM modifié incluant une procédure de Backfitting. Nous avons évalué notre modèle par des simulations et des applications sur des données réelles. Ce modèle nous permet d'étudier l'évolution dans le temps des coefficients associant la qualité de vie des patients aux autres composantes de la douleur chronique pour chaque groupe latent. Le troisième et dernier objectif de la thèse est de proposer un mélange de modèles d'analyse factorielle longitudinale qui permet de résumer plusieurs indicateurs longitudinaux en une ou plusieurs variables latentes qui diffèrent entre les composantes du mélange. Un algorithme EM a été proposé pour estimer le modèle. Des méthodes pour garantir la comparabilité des facteurs latents ont été discuté. L'objectif de ce modèle est d'extraire des indicateurs latents et multidimensionnels qui dépendent à la fois du moment de la mesure dans le parcours du patient et des caractéristiques intrinsèques des patients. Le but ultime étant de faire évoluer l'évaluation de la douleur vers une évaluation multidimensionnelle, variable dans le temps et spécifique au patient, en fonction de ses caractéristiques biopsychosociales.

Mots clés : données longitudinales, douleur chronique, évaluation de la douleur, évaluation multidimensionnelle, analyse factorielle, modèles à coefficients variables, mélange de modèles, modèle à effets mixtes, apprentissage automatique, apprentissage non-supervisé, statistiques non-paramétriques, modèle à classes latentes.

Mixture of random time-varying coefficients and longitudinal factor analysis models and their application to chronic pain multidimensional assessment

Abstract

Multivariate longitudinal data play an essential role in various fields of research, including the medical and behavioral sciences. They allow researchers to test multiple hypotheses, such as the identification of variation over time in a clinical variable. The robustness of the inferences obtained from the analysis of longitudinal data leads to an increasing complexity of statistical processing methods. This complexity arises from the structure of longitudinal data and their statistical properties including intra-individual time dependence. Another degree of complexity may arise from potential heterogeneity due to the existence of several latent subpopulations in the data. Currently, the literature recognizes that pain has a complex longitudinal aspect. It has also been shown that the correlations between the different dimensions of pain become stronger over time. This may be due to the long-term accumulation of biopsychosocial burdens generated by chronic pain and to alterations in brain mechanisms linking pain to emotion. Moreover, the socio-cultural and emotional character of chronic pain makes its evaluation complex since each patient is impacted by the sensory, emotional and functional component differently. The first objective of this thesis is to demonstrate empirically the existence of different latent groups of chronic pain patients through the application of a mixture of mixed-effects model. We show that the quality of life is impacted differently by the sensory, emotional and functional components of pain for each latent group. This provides a rationale for the need for a heterogeneous, patient-specific assessment. The second objective is to propose a new non-parametric model formulated as a time-varying mixture of time-varying coefficient models including random effects processes. We propose a modified EM algorithm to estimate the model parameters. This model allows us to study the evolution over time of the coefficients associating patient quality of life with the other components of chronic pain for each latent group. The third and final objective of the thesis is to propose a new factorial model to analyze heterogeneous multivariate longitudinal data. The proposed model is a mixture of longitudinal factor analysis models that allows summarizing several longitudinal indicators into one or more latent variables, which differ between the mixture components. An EM algorithm has been proposed to estimate the model. Methods to ensure the comparability of the latent factors were discussed. The goal of this model is to extract latent and multidimensional indicators that depend on both the time of measurement in the patient's pathway and the intrinsic characteristics of the patients. The ultimate goal being to shift the assessment of pain to a multidimensional, time-varying and patient-specific assessment based on the patient's biopsychosocial characteristics.

Keywords: longitudinal data, chronic pain, pain assessment, multidimensional assessment, factor analysis, varying-coefficient models, mixture of models, mixed effects model, machine learning, unsupervised learning, Non-parametric statistics, latent class models.