

## 2021-04275 - Post-Doctorant F/H PATH (Parcours paTient en milieu Hospitalier)

Type de contrat : CDD

Niveau de diplôme exigé : Thèse ou équivalent

Fonction : Post-Doctorant

### A propos du centre ou de la direction fonctionnelle

Le centre de recherche Inria Lille – Nord Europe, créé en 2008, compte 360 personnes dont 305 scientifiques répartis dans 15 équipes de recherche. Reconnu pour son implication forte dans le développement socio-économique sur le territoire des Hauts-De-France, le centre de recherche Inria Lille – Nord Europe poursuit une démarche de proximité avec les grandes entreprises et les PME. En favorisant ainsi les synergies entre chercheurs et industriels, Inria participe au transfert de compétences et d'expertises dans les technologies numériques et donne accès aux meilleures recherches européennes et internationales au bénéfice de l'innovation et des entreprises notamment en région.

Depuis plus de 10 ans, le centre Inria Lille – Nord Europe est installé au coeur de l'écosystème universitaire et scientifique lillois ainsi qu'au coeur de la Frenchtech avec un showroom technologique, basé avenue de Bretagne à Lille, sur le site d'excellence économique consacré aux technologies de l'information et de la communication (TIC) qu'est EuraTechnologies

### Contexte et atouts du poste

Le projet de recherche s'inscrit dans le cadre d'une action exploratoire de l'INRIA d'un consortium de médecins, bio-statisticiens, statisticiens. L'objectif est de permettre une meilleure compréhension des étapes clés dans les parcours de soins des patients en associant les producteurs de données au plus proche du patient, ceux qui les gèrent, ceux qui les pré-traitent, ceux qui les analysent, pour avoir au final un résultat au plus proche du terrain et un retour vers le clinicien et le patient le plus efficient possible.

Le projet, par essence interdisciplinaire et exploratoire, est dans la continuité des collaborations passées entre les membres des deux unités INRIA-MODAL et METRICS (Université de Lille). Il ne pourrait pas être mené sans une étroite collaboration entre les médecins et les chercheurs en mathématiques appliquées.

INRIA-MODAL : Sophie Dabo, Guillemette Marot, Vincent Vandewalle, Cristian Preda et Christophe Biernacki

METRICS-Université de Lille : Evgeniya Babykina, Jean-Baptiste Beuscart, Emmanuel Chazard, Cyrielle Dumont, Grégoire Ficheur, Michaël Génin, Antoine Lamer

Les systèmes de santé européens sont confrontés à de multiples défis dont le vieillissement de la population, l'augmentation des maladies chroniques et des patients souffrant de multi-morbidité, des ressources financières et humaines contraintes [4]. La réponse à ces défis repose notamment sur une organisation des soins en parcours de soins, justifiée par une littérature scientifique abondante [20] et soutenue en France par les orientations politiques régionales et nationales. Selon la Haute Autorité de Santé (HAS), le parcours de soins ne s'entend pas simplement comme une succession d'actes ponctuels et indépendants des producteurs de soins, mais comme "le juste enchaînement et au bon moment de ces différentes compétences professionnelles liées directement ou indirectement aux soins [...]". Plus généralement, le parcours de soins, comme défini dans [18], est l'intervention complexe pour la prise de décision et l'organisation de processus de soins pour un groupe de patients bien défini pendant une période bien définie.

L'analyse des parcours de soins et leur adéquation aux besoins et aux moyens est ainsi devenue un enjeu majeur sur les plans scientifique et administratif. Si les données numériques disponibles à cet effet augmentent rapidement, les méthodes et les outils statistiques mis à disposition des chercheurs et des tutelles sanitaires restent limités, et peu efficaces.

Les types de parcours de soins sont très nombreux. Dans le cadre de cette action exploratoire, nous proposons de nous intéresser à deux cas d'application : 1) un parcours de soins ambulatoire (lien ville-hôpital) ; 2) un parcours de soins intra-hospitalier. Ce choix se justifie par une expertise solide de METRICS de ces parcours, ancrée sur plusieurs années de recherche, ainsi que des liens étroits avec les cliniciens experts de ces questions.

L'informatisation des fournisseurs de soins (hôpitaux, cabinets médicaux, laboratoires de biologie médicale) et des assurances (dont l'Assurance Maladie) a permis l'accumulation de données massives relatives au soin des patients déconcentrées [5]. Ces données peuvent être réutilisées (data reuse) [6] afin d'étudier les parcours de soins des patients, mais aussi pour développer des modèles prédictifs de ces parcours, susceptibles d'être intégrés dans des procédures d'intelligence artificielle (IA). Ces données sont :

- toutes temps-dépendantes ;
- fortement hétérogènes (actes et diagnostics codés, résultats d'analyses de biologie médicale, médicaments administrés, données de parcours, etc.) ;
- qualitatives à plusieurs milliers de modalités possibles, et donc très fortement déséquilibrées ;
- principalement constituées de données manquantes, et ce presque jamais par hasard ;
- représentées dans des dizaines ou centaines de tables en relation ;
- souvent spatiales.

En l'état actuel des méthodologies, l'extraction de données et de leurs caractéristiques paraît inaccessible à la plupart des équipes hors du domaine de la santé [13]. L'association de cliniciens experts des parcours considérés, d'experts en données médicales, de statisticiens spécialisés devrait nous permettre d'automatiser, au moins en partie, certaines étapes d'extraction et de lever des verrous méthodologiques dans la modélisation de ce type de données.

### Informations générales

- **Thème/Domaine** : Optimisation, apprentissage et méthodes statistiques Statistiques (Big data) (BAP E)
- **Ville** : Villeneuve d'Ascq
- **Centre Inria** : CRI Lille - Nord Europe
- **Date de prise de fonction souhaitée** : 2022-03-01
- **Durée de contrat** : 1 an, 6 mois

### Contacts

- **Equipe Inria** : MODAL
- **Recruteur** :  
Dabo Sophie / Sophie.Dabo@inria.fr

### A propos d'Inria

Inria est l'institut national de recherche dédié aux sciences et technologies du numérique. Il emploie 2600 personnes. Ses 200 équipes-projets agiles, en général communes avec des partenaires académiques, impliquent plus de 3500 scientifiques pour relever les défis du numérique, souvent à l'interface d'autres disciplines. L'institut fait appel à de nombreux talents dans plus d'une quarantaine de métiers différents. 900 personnels d'appui à la recherche et à l'innovation contribuent à faire émerger et grandir des projets scientifiques ou entrepreneuriaux qui impactent le monde. Inria travaille avec de nombreuses entreprises et a accompagné la création de plus de 180 start-up. L'institut s'efforce ainsi de répondre aux enjeux de la transformation numérique de la science, de la société et de l'économie.

### L'essentiel pour réussir

**Compétences techniques et niveau requis** : Expérience en modélisation statistique (des connaissances en analyse de données temps-dépendantes)

### Consignes pour postuler

CV + lettre de motivation + lettre(s) de recommandation

#### Sécurité défense :

Ce poste est susceptible d'être affecté dans une zone à régime restrictif (ZRR), telle que définie dans le décret n°2011-1425 relatif à la protection du potentiel scientifique et technique de la nation (PPST). L'autorisation d'accès à une zone est délivrée par le chef d'établissement, après avis ministériel favorable, tel que défini dans l'arrêté du 03 juillet 2012, relatif à la PPST. Un avis ministériel défavorable pour un poste affecté dans une ZRR aurait pour conséquence l'annulation du recrutement.

#### Politique de recrutement :

Dans le cadre de sa politique diversité, tous les postes Inria sont accessibles aux personnes en situation de handicap.

**Attention**: Les candidatures doivent être déposées en ligne sur le site Inria. Le traitement des candidatures adressées par d'autres canaux n'est pas garanti.

La construction des parcours de soin à partir des données brutes implique une décision experte (médicale) et la mise en œuvre de processus automatisés. Le détail des décisions expertes n'est jamais explicité lors des publications, ce qui freine la reproductibilité des travaux et, à notre connaissance, ne fait pas l'objet d'un consensus méthodologique. Les processus automatisés que nous proposons d'utiliser reposent sur des algorithmes d'analyse statistique (clustering, analyse de classes latentes, embedding) au cœur du projet scientifique de MODAL.

Une fois les données nécessaires pour la construction d'un parcours de soins acquises, plusieurs problèmes apparaissent dans l'exploration et l'analyse de ces données : (1) des populations hétérogènes; (2) le critère de jugement perturbé par des événements concurrents, comme le décès; (3) le critère de jugement mesuré comme une donnée répétée (par exemple, l'évolution d'un score fonctionnel) dont l'évolution n'est pas nécessairement homogène au cours du temps; (4) la réponse au traitement évaluée à partir de plusieurs critères de jugement simultanément (outcome multivarié), comme par exemple, temps jusqu'au décès, évolution de la qualité de vie et de la cognition; (5) la prise en compte de facteurs d'exposition longitudinaux (ex : mesure journalière de pollution de l'air par des capteurs sur un territoire) et évaluation de leur impact sur les variations spatiales de pathologies ou détection de clusters spatiaux d'événements temporellement récurrents (ex : zones géographiques présentant un taux anormalement élevé de ré-hospitalisations suite à une chirurgie, etc.); description de la dynamique temporelle de variations spatiales d'événements de santé dans le cadre de la recherche de signaux étiologiques; (6) la mesure à différents temps de milliers de variables simultanément, structurées ou non par bloc, par exemple des données omiques.

Des déplacements réguliers sont prévus pour ce poste entre INRIA, METRICS (Université de Lille) et le CHU de Lille.

## Mission confiée

Bien que de nombreuses approches statistiques (clustering, régression, analyse de survie) de données complexes avec dépendance temporelle ou spatio-temporelle ont connu un essor important cette dernière décennie ([8, 12, 1, 10], etc.), elles nécessitent d'être étendues aux données de parcours patient des deux cas d'usages décrits précédemment, afin de répondre aux questions suivantes :

1. Identifier des parcours-type et des parcours atypiques.
2. Prédire des états futurs d'un parcours de soin.
3. Prédire des événements (certains récurrents) : ré-hospitalisations, décès, interventions.

Afin d'analyser les deux cas d'application proposés (parcours de soins ambulatoire intra-hospitalier) et répondre aux questions ci-dessus, de récents travaux sur les parcours patients sur la visualisation [11] (METRICS), sur des algorithmes d'analyse de séquences [16, 3] (ORPAILLEUR, LACODAM), et sur l'analyse des risques de ré-hospitalisation [21] (METRICS) seront exploités dans un premier temps, puis étendus à l'aide des recherches de MODAL, METRICS et de la littérature sur des :

- modèles joints, temporels, spatio-temporels
- modèles génératifs de parcours patient dans le même esprit que les modèles conjoints
- modèles d'apprentissage supervisé avec réponse fonctionnelle multivariée et modèles de séries temporelles fonctionnelles avec des outliers/extrêmes développés dans la littérature.

L'implémentation de ces modèles existants, leurs extensions, applications et interprétations au domaine clinique nécessitent des compétences cliniques, statistiques et en optimisation numérique.

Les missions confiées concernent donc la production/le développement des méthodes statistiques pour la construction/analyse du parcours patient à travers des cas applications portant sur la ré-hospitalisation de la personne âgée et les complications postopératoires.

### Références :

[1] Akim Adekpedjou and Sophie Dabo-Niang. Semiparametric estimation with spatially correlated recurrent events. *Scandinavian Journal of Statistics*, 2020.

[2] Shola Adeyemi, Eren Demir, and Thierry Chausalet. Towards an evidence-based decision making health care system management : Modelling patient pathways to improve clinical outcomes. *Decision Support Systems*, 55(1) :117–125, 2013.

[3] Johanne Bakalara, Thomas Guyet, Olivier Dameron, André Happe, and Emmanuel Oger. An extension of chronicles temporal model with taxonomies - Application to epidemiological studies. In *HEALTHINF 2021 - 14th International Conference on Health Informatics*, pages 1–10, online, France, February 2021.

[4] Karen Barnett, Stewart W Mercer, Michael Norbury, Graham Watt, Sally Wyke, and Bruce Guthrie. Epidemiology of multimorbidity and implications for health care, research, and medical education : a cross-sectional study. *The Lancet*, 380(9836) :37–43, 2012.

[5] Emilie Baro, Samuel Degoul, Régis Beuscart, and Emmanuel Chazard. Toward a literature-driven definition of big data in healthcare. *BioMed research international*, 2015, 2015.

[6] Emmanuel Chazard, Grégoire Ficheur, Alexandre Caron, Antoine Lamer, Julien Labreuche, Marc Cuggia, Michaël Genin, Guillaume Bouzille, Alain Duhamel, et al. Secondary use of healthcare structured data : The challenge of domain-knowledge based extraction of features. *EFMI-STC*, pages 15–19, 2018.

[7] Elias Egho, Nicolas Jay, Chedy Raissi, Gilles Nuemi, Catherine Quantin, and Amedeo Napoli. An approach for mining care trajectories for chronic diseases. In *AIME 2013 - 14th Conference on Artificial Intelligence in Medicine*, volume 7885, Murcia, Spain, May 2013. Springer.

[8] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3) :37–37, 1996.

[9] Kelsey Flott, Ara Darzi, and Erik Mayer. Care pathway and organisational features driving patient experience : statistical analysis of large nhs datasets. *BMJ open*, 8(7) :e020411, 2018.

[10] Agnieszka Krol, Audrey Mauguen, Yassin Mazroui, Alexandre Laurent, Stefan Michiels, and Virginie Rondeau. Tutorial in joint modeling and prediction : a statistical software for correlated longitudinal outcomes, recurrent events and a terminal event. *arXiv preprint arXiv:1701.03675*, 2017.

[11] Antoine Lamer, Gery Laurent, Sylvia Pelayo, Mehdi El Amrani, Emmanuel Chazard, and Romaric Marcilly. Exploring patient path through sankey diagram : a proof of concept. *Studies in health technology and informatics*, 270 :218–222, 2020.

[12] Yi Li and Xihong Lin. Semiparametric normal transformation models for spatially correlated survival data. *Journal of the American Statistical Association*, 101(474) :591–603, 2006.

[13] Stephane M Meystre, Christian Lovis, Thomas Bürkle, Gabriella Tognola, Andrius Budrionis, and Christoph U Lehmann. Clinical data reuse or secondary use : current status and potential future

progress. Yearbook of medical informatics, 26(1) :38, 2017.

[14] Rupert M Pearse, Rui P Moreno, Peter Bauer, Paolo Pelosi, Philipp Metnitz, Claudia Spies, Benoit Vallet, Jean-Louis Vincent, Andreas Hoeft, Andrew Rhodes, et al. Mortality after surgery in europe : a 7 day cohort study. The Lancet, 380(9847) :1059–1065, 2012.

[15] Adam Perer, Fei Wang, and Jianying Hu. Mining and exploring care pathways from electronic medical records with visual analytics. Journal of biomedical informatics, 56 :369–378, 2015.

[16] Gabin Personeni, Marie-Dominique Devignes, Michel Dumontier, Malika Smail-Tabbone, and Adrien Coulet. Extraction d'association d'EIM a` partir de dossiers patients : expérimentation avec les structures de patrons et les ontologies. In Deuxième Atelier sur l'Intelligence Artificielle et la Santé, Atelier IA & Santé, Montpellier, France, June 2016.

[17] Cristian Preda, Quentin Grimonprez, and Vincent Vandewalle. cfda : an r package for categorical functional data analysis. 2020.

[18] Guus Schrijvers, Arjan van Hoorn, and Nicolette Huiskes. The care pathway : concepts and theories : an introduction. International journal of integrated care, 12(Special Edition Integrated Care Pathways), 2012.

[19] Antonia E Stephen and David L Berger. Shortened length of stay and hospital cost reduction with implementation of an accelerated clinical care pathway after elective colon resection. Surgery, 133(3) :277–282, 2003.

[20] Diane E Threapleton, Roger Y Chung, Samuel YS Wong, Eliza Wong, Patsy Chau, Jean Woo, Vincent CH Chung, and Eng-Kiong Yeoh. Integrated care for older populations and its implementation facilitators and barriers : A rapid scoping review. International Journal for Quality in Health Care, 29(3) :327–334, 2017.

[21] Fabien Visade, Genia Babykina, Antoine Lamer, Marguerite-Marie Defebvre, David Verloop, Grégoire Ficheur, Michael Genin, François Puisieux, and Jean-Baptiste Beuscart. Importance of previous hospital stays on the risk of hospital re-admission in older adults : a real-life analysis of the paerpa study population. Age and Ageing, 50(1) :141–146, 2021.

[22] Krist Wongsuphasawat and David Gotz. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. IEEE Transactions on Visualization and Computer Graphics, 18(12) :2659– 2668, 2012.

## Principales activités

**Principales activités :**

La personne recrutée s'attachera à faire le lien entre l'extraction de features (faite en particulier par un ingénieur) et l'analyse de données (apprentissage statistique de données temporelles, spatio-temporelles), recenser les méthodes d'analyse statistiques pertinentes, identifier les limites des méthodes statistiques actuelles, proposer des développements au regard des objectifs ci-dessus.

**Activités complémentaires :**

Présenter les travaux aux différents partenaires, dans des manifestations scientifiques

Analyser les demandes des différents partenaires

## Compétences

Compétences techniques et niveau requis : Développement de packages R, Python, niveau expert.

Langues : Français, Anglais

Compétences additionnelles appréciées : Autonomie, Rigueur, Passion de l'innovation, des applications

## Avantages

- Restauration subventionnée
- Transports publics remboursés partiellement
- Congés: 7 semaines de congés annuels + 10 jours de RTT (base temps plein) + possibilité d'autorisations d'absence exceptionnelle (ex : enfants malades, déménagement)
- Possibilité de télétravail (après 6 mois d'ancienneté) et aménagement du temps de travail
- Équipements professionnels à disposition (visioconférence, prêts de matériels informatiques, etc.)
- Prestations sociales, culturelles et sportives (Association de gestion des œuvres sociales d'Inria)
- Accès à la formation professionnelle
- Sécurité sociale

## Rémunération

2 653€ brut mensuel