



**Offer #2022-05097**

## **Engineer PATH (PATient PaThway in the Hospital environment) (F/M)**

**Contract type :** Fixed-term contract

**Level of qualifications required :** Graduate degree or equivalent

**Fonction :** Temporary scientific engineer

**Level of experience :** Up to 3 years

### **About the research centre or Inria department**

The Inria Lille - Nord Europe research center, created in 2008, employs 360 people including 305 scientists in 15 research teams. Recognized for its strong involvement in the socio-economic development of the Hauts-De-France region, the Inria Lille - Nord Europe research center pursues a close relationship with large companies and SMEs. By promoting synergies between researchers and industrialists, Inria participates in the transfer of skills and expertise in digital technologies and provides access to the best European and international research for the benefit of innovation and companies, particularly in the region.

For more than 10 years, the Inria Lille - Nord Europe center has been located at the heart of the university and scientific ecosystem in Lille, as well as at the heart of Frenchtech, with a technology showroom based on avenue de Bretagne in Lille, on the site of economic excellence dedicated to information and communication technologies (ICT) that is EuraTechnologies.

### **Context**

The research project is part of an INRIA exploratory action of a consortium of physicians, bio-statisticians, statisticians. The objective is to allow a better understanding of the key stages in the care of patients by associating the producers of data as close as possible to the patient, those who manage them, those who pre-process them, those who analyze them, in order to have a result as close to the field as possible and a return to the clinician and the patient as efficient as possible.

The project, which is essentially interdisciplinary and exploratory, is in line with past collaborations between members of the two units INRIA-MODAL and METRICS (University of Lille). It could not be conducted without a close collaboration between physicians and researchers in applied mathematics.

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

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

European healthcare systems are faced with multiple challenges, including an aging population, an increase in chronic diseases and patients with multi-morbidity, and limited financial and human resources [4]. The response to these challenges is based in particular on the organization of care into care pathways, justified by abundant scientific literature [20] and supported in France by regional and national political orientations. According to the French National Authority for Health (HAS), the care pathway is not simply a succession of one-off acts independent of the producers of care, but "the right sequence and timing of these different professional skills linked directly or indirectly to care [...]". More generally, the care pathway, as defined in [18], is the complex intervention for decision-making and organization of care processes for a well-defined group of patients during a well-defined period of time.

The analysis of care pathways and their adequacy to needs and means has thus become a major scientific and administrative challenge. Although the numerical data available for this purpose are increasing rapidly, the statistical methods and tools available to researchers and health authorities remain limited and inefficient.

The types of care pathways are very numerous. Within the framework of this exploratory action, we propose to focus on two cases of application: 1) an ambulatory care pathway (city-hospital link); 2) an intra-hospital care pathway. This choice is justified by METRICS' solid expertise in these pathways, based on several years of research, as well as close links with clinicians who are experts in these issues.

The computerization of health care providers (hospitals, medical practices, medical laboratories) and insurance companies (including the French National Health Insurance) has led to the accumulation of

massive data related to the care of de concentrated patients [5]. These data can be reused (data reuse) [6] in order to study patient care pathways, but also to develop predictive models of these pathways, which can be integrated into artificial intelligence (AI) procedures. These data are :

- all time-dependent;
- highly heterogeneous (coded procedures and diagnoses, results of medical tests, drugs administered, pathway data, etc.)
- qualitative, with several thousand possible modalities, and therefore very strongly unbalanced
- mainly made up of missing data, and almost never by chance;
- represented in dozens or hundreds of related tables;
- often spatial.

In the current state of methodologies, the extraction of data and their characteristics seems inaccessible to most teams outside the health field [13]. The association of clinicians who are experts in the pathways considered, experts in medical data, and specialized statisticians should enable us to automate, at least in part, certain extraction steps and to remove methodological obstacles in the modeling of this type of data.

The construction of care pathways from raw data implies an expert (medical) decision and the implementation of automated processes. The details of the expert decisions are never explained in the publications, which hinders the reproducibility of the work and, to our knowledge, is not the subject of a methodological consensus. The automated processes that we propose to use are based on statistical analysis algorithms (clustering, latent class analysis, embedding) at the heart of the MODAL scientific project.

Once the data necessary for the construction of a care pathway are acquired, several problems appear in the exploration and analysis of these data: (1) heterogeneous populations; (2) the endpoint disrupted by competing events, such as death; (3) the endpoint measured as repeated data (e.g., change in a functional score) whose evolution is not necessarily homogeneous over time; (4) the response to the treatment evaluated from several criteria of judgment simultaneously (multivariate outcome), such as for example, time to death, evolution of the quality of life and cognition; (5) the taking into account of longitudinal exposure factors (ex: (5) taking into account longitudinal exposure factors (e.g. daily measurement of air pollution by sensors on a territory) and evaluating their impact on spatial variations of pathologies or detection of spatial clusters of temporally recurrent events (e.g. geographical areas with an abnormally high rate of re-hospitalization following surgery, etc.); description of the temporal dynamics of the pathologies (e.g. time to death, quality of life, cognition, etc.); and (6) taking into account the impact of longitudinal exposure factors (e.g. time to death, quality of life, etc.); description of the temporal dynamics of spatial variations of health events in the context of the search for etiological signals; (6) the measurement at different times of thousands of variables simultaneously, structured or not by block, for example omics data.

Regular travel is expected for this position between INRIA, METRICS (University of Lille) and the University Hospital of Lille.

## Assignment

Although many statistical approaches (clustering, regression, survival analysis) of complex data with temporal or spatio-temporal dependence have been developed in the last decade ([8, 12, 1, 10], etc.), they need to be extended to patient pathway data of the two use cases described above, in order to answer the following questions:

1. Identify typical and atypical pathways.
2. Predict future states of a care pathway.
3. Predicting events (some recurrent): re-hospitalisations, deaths, interventions.

In order to analyse the two proposed application cases and answer the above questions, recent work on patient pathways and visualisation [11] (METRICS), on sequence analysis algorithms [16, 3] (ORPAILLEUR, LACODAM), and on the analysis of the risks of re-hospitalisation [21] (METRICS) will be exploited in the first instance, and then extended with the help of MODAL and METRICS research and the literature on :

- joint, temporal, spatio-temporal models
- generative patient pathway models in the same spirit as the joint models
- supervised learning models with multivariate functional response and functional time series models with outliers/extremes developed in the literature.

The implementation of these existing models, their extensions, applications and interpretations in the clinical field require clinical, statistical and numerical optimisation skills.

## References

- [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 Chausaulet. Towards an evidence-based decision making heal-

thcare system management: Modelling patient pathways to improve clinical outcomes. *Decision Support Systems*, 55(1) :117-125, 2013.

[3] Johanne Bakalara, Thomas Guyet, Olivier Dameron, Andr e 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, R gis Beuscart, and Emmanuel Chazard. Toward a literature-driven definition of big data in healthcare. *BioMed research international*, 2015, 2015.

[6] Emmanuel Chazard, Gr goire Ficheur, Alexandre Caron, Antoine Lamer, Julien Labreuche, Marc Cuggia, Micha el 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 Ra issi, 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 Kr  ol, Audrey Mauguen, Yassin Mazroui, Alexandre Laurent, Stefan Michiels, and Virginie Ron-deau. 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 Bu rkle, 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 Pearce, 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 Sma  il-Tabbone, and Adrien Cou-let. ADR association extraction from patient records: exp erimentation with pattern structures and ontologies. In *Deuxi eme Atelier sur l'Intelligence Artificielle et la Sant e*, Atelier IA & Sant e, Montpellier, France, June 2016.

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

6

[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, Gr goire Ficheur, Micha el Genin, Franc ois 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.

## Main activities

### Main activities

The person recruited will work closely with clinicians and their various partners for data extraction and features and with a postdoc in Statistics. He/she will build databases from different sources stored at the University Hospital of Lille, clean but not usable for statistical analysis, he/she will perform preliminary statistical analyses. The raw data is not in the format of the target object (patient pathway). The target object is complex, it is a pathway that has an initial time  $T_0$ , a terminal time  $T_{fin}$ , intermediate stages, characteristics that can vary according to the stages, possibilities of combinations of stages (filled or not). The construction of this target object (according to the use case considered) is a research work (contrary to that of the engineer sought for this position), carried out by the project researchers and the postdoc recruited.

Weekly meetings will be scheduled at the Lille University Hospital or at INRIA. The person recruited will be based mainly for the first 6 months at the Lille University Hospital to carry out the extraction of the data available on this site, the rest of the time will be shared between the INRIA centre in Lille and the University Hospital.

The production of the patient pathway extraction software and the technical report at the end of the contract would be the ultimate goal.

### Additional activities

Present the work to different partners, in scientific events Analyze the requests of different partners

## Skills

**Technical skills and level required:** Development of R packages, Python, expert level.

**Languages:** French, English

**Additional skills appreciated:** Autonomy, Rigor, Passion for innovation, applications

## Benefits package

- Subsidized meals
- Partial reimbursement of public transport costs
- Leave: 7 weeks of annual leave + 10 extra days off due to RTT (statutory reduction in working hours) + possibility of exceptional leave (sick children, moving home, etc.)
- Possibility of teleworking and flexible organization of working hours
- Professional equipment available (videoconferencing, loan of computer equipment, etc.)
- Social, cultural and sports events and activities
- Access to vocational training
- Social security coverage

## Remuneration

Gross salary: According to the profile

## General Information

- **Theme/Domain :** Optimization, machine learning and statistical methods Statistics (Big data) (BAP E)
- **Town/city :** Villeneuve d'Ascq
- **Inria Center :** [Centre Inria de l'Université de Lille](#)
- **Starting date :** 2023-09-01
- **Duration of contract :** 1 year
- **Deadline to apply :** 2023-09-30

## Contacts

- **Inria Team :** [MODAL](#)
- **Recruiter :**  
Dabo Sophie / [Sophie.Dabo@inria.fr](mailto:Sophie.Dabo@inria.fr)

## About Inria

Inria is the French national research institute dedicated to digital science and technology. It employs 2,600 people. Its 200 agile project teams, generally run jointly with academic partners, include more than 3,500 scientists and engineers working to meet the challenges of digital technology, often at the

interface with other disciplines. The Institute also employs numerous talents in over forty different professions. 900 research support staff contribute to the preparation and development of scientific and entrepreneurial projects that have a worldwide impact.

## The keys to success

Experience in statistics, data science, mastery of R, Python, Excel, HTML/CSS, MySQL. A deep interest in applications and software development is highly recommended.

**Warning :** you must enter your e-mail address in order to save your application to Inria. Applications must be submitted online on the Inria website. Processing of applications sent from other channels is not guaranteed.

## Instruction to apply

### **Defence Security :**

This position is likely to be situated in a restricted area (ZRR), as defined in Decree No. 2011-1425 relating to the protection of national scientific and technical potential (PPST). Authorisation to enter an area is granted by the director of the unit, following a favourable Ministerial decision, as defined in the decree of 3 July 2012 relating to the PPST. An unfavourable Ministerial decision in respect of a position situated in a ZRR would result in the cancellation of the appointment.

### **Recruitment Policy :**

As part of its diversity policy, all Inria positions are accessible to people with disabilities.