can be integrated into artificial intelligence (AI) procedures. These data are: patient care pathways, but also to develop predictive models of these pathways, which Insurance) has led to the accumulation of massive data related to the care of laboratories) and insurance companies (including the French National Health The computerization of health care providers (hospitals, medical practices, medical research, as well as close links with clinicians who are experts in these issues. justifies METRICS' solid expertise in these pathways, based on several years of exploratory action, we propose to focus on two cases of application: 1) an ambulatory The types of care pathways are very numerous. Within the framework of this research, 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 EurTechnologies. 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, 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;

A propos du centre ou de la direction fonctionnelle

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 EurTechnologies.

Contexte et atouts du poste

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, Guillaume Marot, Vincent Vanderwalle, Cristian Preda and Christophe Biernacki

METRICS-University of Lille: Evgeniya Babyska, Jean-Baptiste Beuscart, Emmanuel Chazard, Cyrielle Dumont, Grégoire Fichere, Michael Gnin, 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 means and needs 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, 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:
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 abnormal 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.

Mission confiée

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.
3. Predicting events (some recurrent): re-hospitalisations, deaths, interventions.

In order to analyze 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-hospitalization [2] (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


Gabin Personeni, Marie-Dominique Devignes, Michel Dumontier, Malika Sma-Dil-Talbione, and Adrien Coulet. ADR association extraction from patient records: experimentation with pattern structures and ontologies. In Deuxième Atelier sur l'Intelligence Artificielle et la Santé, Atelier IA Santé, Montpellier, France, June 2016.


**Principales activités**

**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 T0, a terminal time Tfin, 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

**Compétences**

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

**Languages**: French, English

**Additional skills appreciated**: Autonomy, Riger, Passion for innovation, applications
Avantages

- 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

Rémunération

Gross salary: According to the profile