aviesan
alliance nationale
pour les sciences de la vie et de la santé

ITMO Cancer

Instituts thématiques

Inserm
Institut national
de la santé et de la recherche médicale

# APPEL À PROJETS 2021
# 2021 CALL FOR PROPOSALS

## Approches interdisciplinaires des processus oncogéniques et perspectives thérapeutiques : Apports à l'oncologie des mathématiques et de l'informatique

## Interdisciplinary approaches in oncogenic processes and therapeutic perspectives: Contributions of mathematics and informatics to oncology

**DOSSIER DE CANDIDATURE**
**EN ANGLAIS OBLIGATOIREMENT**

**APPLICATION FILE**
**STRICTLY IN ENGLISH**

Le dossier complet doit être déposé par le coordonnateur sur le site internet EVA3 (https://eva3-accueil.inserm.fr/sites/eva/appels-a-projets/pca/Pages/MIC.aspx) avant 17h le 25 Février 2021.

The full application file must be submitted by the coordinator on the EVA3 website (https://eva3-accueil.inserm.fr/sites/eva/appels-a-projets/pca/Pages/MIC.aspx) before 17h of 25th February 2021.

# TABLE OF CONTENTS

## Part I

### 1-1 *Registry office of the principal investigator*

| | |
|---|---|
| Nom du Coordonnateur du projet / *Last name of the Project Coordinator* | SAVY |
| Prénom du Coordonnateur du projet / *First name of the Project Coordinator* | Nicolas |
| Laboratoire de rattachement / *Name of the research laboratory* | Institut de Mathématiques de Toulouse |
| Directeur du laboratoire / *Head of the laboratory* | Franck BARTHE |
| Responsable d'équipe / *Team leader* | Pierre NEUVIAL |
| Adresse du laboratoire / *Laboratory postal address* | 118, route de Narbonne |
| Ville / *City* | Toulouse |

### 1-2 *Project ID*

| | |
|---|---|
| Titre du projet en Anglais/ Project title in English | Machine Learning for Healthcare Pathways Analysis |
| Durée prévue du projet (mois) / *Scheduled duration of the project (month)* | 36 months |
| Nombre d'équipes participantes / *Number of associated teams* | 3 |
| Montant total demandé à l'Inserm (toutes équipes confondues) / *Total amount requested from Inserm (all teams taken as a whole)* | **204 776 euros** |

### 1-3 *Relevance of the project with the objectives of the call for proposals*

| Approches mathématiques et informatiques / *Mathematical and computer approaches* | |
|---|---|
| ☒ | **Modélisation (développement et validation)** */ Modeling (development and validation)* |
| ☒ | **Simulation** */ Simulation* |
| ☒ | **Algorithmes et méthodes computationnelles** */ Computational algorithms and methods* |
| ☒ | **Intégration, visualisation, représentation, fouille et analyse de données hétérogènes de grandes dimensions** */ Integration, visualization, representation, search and analysis of large heterogeneous data* |
| ☒ | **Développement de méthodes d'apprentissage** */ Development of learning methods* |
| ☒ | **Analyse de réseaux (dont analyse causale, …)** */ Network analysis (including causal analysis ...)* |

| Champs d'études / Domain | |
|---|---|
| ☐ | **Processus oncogéniques (histoire naturelle des cancers, ontologie tumorale, évolution clonale, croissance tumorale, développement métastatique, ...)** / *Oncogenic processes (natural history of cancers, tumor ontology, clonal evolution, tumor growth, metastatic development, ...)* |
| ☐ | **Relation de la tumeur au microenvironnement** / *Relationship of the tumor with the microenvironment* |
| ☐ | **Rôle du système immunitaire dans le développement tumoral** / *Role of the immune system in tumor development* |
| ☐ | **Réponses et résistance aux traitements, rechute** / *Response and resistance to treatment, relapse* |
| ☒ | **Aides au diagnostic, prognostic, suivi thérapeutique** / *Diagnostic aids, prognosis, therapeutic follow-up* |
| ☒ | **Drug design, optimisation/combinaison de traitements** */ Drug design, optimization / combination of treatments* |

### 1-4 *Abstract*

**Scientific background**. The issue of healthcare pathways is an important one, particularly in the context of chronic diseases such as cancer. At the collective level, with the classification and study of trajectories, as well as at the individual level with the search for optimal and personalized decision rules, many methods exist. These methods have shown their performances but also their weaknesses which make their use constrained to stringent hypotheses or contexts.

**Project description**. The ML4PHA project proposes to investigate the relevance of using certain innovations in learning in this context. The project is divided into two axes. The first axis "collective scale" will study the relevance of the use of Dynamic Bayesian Networks for the modeling of care pathways on the one hand and the use of agent-based models for the simulation of healthcare pathways on the other hand. This research will be applied to data on costs associated with thyroid cancer and then with chronic myeloid leukemia. The second axis "individual scale" will study the relevance of the use of reinforcement learning techniques like deep Q-learning techniques which would make it possible to apprehend more complex situations than the usual methods and which will be applied to the study of therapeutic decision making for chronic myeloid leukemia. We will address the question of the "explorability" of the trajectories using learning techniques allowing to integrate the a priori of experts and explore the relevance of introducing Generative Adversarial Networks for small databases.

**Expected results**. The expected results are of several types: expanding the arsenal of care pathway analysis techniques, showing its applicability to cancer data, providing the community with tools to facilitate the use of these techniques and providing clinicians with decision-making tools.

| **Domaines et mots clés / Topics and keywords** | |
|---|---|
| ITMO | Santé Publique |
| Domaines ITMO / ITMO Domains | Biostatistique et Modélisation statistique<br>Economie de la santé - Systèmes sanitaires |
| Mots clés / Keywords | |
| Healthcare Pathway ; Dynamic Bayesian network ; reinforcement learning ; sequence analysis ; Generative Adversarial Network | |

* *La liste des domaines sont disponibles dans le guide du candidat. The list of domains is available in the "guide du candidat".*

## Part II

### 2-1 Scientific project

#### 2.1.1 Issue, hypothesis and objective(s) of the investigations

The issue of the healthcare pathway (or healthcare trajectory) is a major one, particularly in the context of chronic diseases. Chronic disease is characterized by a sequence of clinical observations or a sequence of care, so the natural study context is indeed a trajectory in order to take into account the information contained in the observation dependency. However, the study of healthcare pathways is complex. It is based on a subtle methodological triptych:

• A definition of the objective of the study and the object of the study (definition of the pathway),
• A relevant analysis technique to achieve this objective,
• Data available to achieve this goal.

The study of healthcare pathway cannot therefore be conceived without an interdisciplinary approach integrating care specialists, data specialists and modeling specialists. This project is part of this process. Around a consortium bringing together these skills, the project aims to study different technical aspects of the analysis of healthcare pathways in the specific field of cancer.

The initiation of this project is based on two studies of healthcare pathways:

- An economic study, led by teams 1 and 2, is currently in progress about the healthcare pathways and cost of Thyroid Cancer (TC) with a good prognosis. Main objectives are 1) to estimate lifetime costs of Thyroid Cancer in comparison with controls using a cox model and 2) to describe the 3 years healthcare pathways after partial or total thyroidectomy using the optimal matching.
- A clinical study, led by teams 1 and 3, addressing leukemia patients' treatment pathways. In this study we seek for an optimal therapeutic strategy, i.e. a combination of treatment lines, based on both the characteristics of the disease and of the patients, regarding the response to treatment and cancer relapse.

These studies have enabled us to highlight the strengths of certain techniques for analyzing healthcare pathways, but also the weaknesses of these techniques and the study of possible improvements to these techniques.

The issue of healthcare pathways is very vast. Indeed, the information contained in a trajectory is polymorphic, at the same time temporal, geographical, demographic, medical, and economic, the objectives and therefore the associated analysis methods are varied.
The scales of analysis are diverse: the scale of the population of analysis for an individual approach (precision medicine) to a collective approach (medico-economic for example), the time scale for a short-term approach (short horizon window) or a whole life approach (long horizon window or by simulation), the scale of the trajectory for a specific approach (dose finding for example) to a holistic approach.

The objectives of the analysis of healthcare pathways are also numerous among which:

- Define "typical" trajectories and study the determinants of these,
- Define trajectories that lead to a given state of health and determine the risk factors of being in a given state at a given time,
- Describe and optimize the modalities of therapeutic management of certain patients,
- Understand and optimize the care of patients with chronic pathology,
- Measure the impact of intervention on the treatment pathway through simulation under given conditions,
- Model and standardize care,
- Evaluate the efficiency of support.

The variety of possible objectives and the variety of analysis scales mean that many techniques for analyzing healthcare pathways have been tested, in particular in recent years with the implementation of methods derived from Artificial Intelligence: unsupervised methods for the classification of care trajectories, supervised methods (Markov model / latent class model) to identify determinants of healthcare trajectories and reinforcement methods (Dynamic Treatment Regimes) to identify optimal care trajectories. As such, oncology is a particularly suitable field of application for investigating many of these methods. While these techniques have shown their relevance in specific contexts, they also show a number of weaknesses that we plan to identify. Alternative or at least complementary methods will also be investigated by specifying their methodological context and by measuring their performance on real data.

Our attention will focus on two fundamental aspects of healthcare pathways in the context of cancer, namely the study of the patient's trajectory with an axis oriented "collective approach" with applications in health economics and an axis "individual approach" with clinical applications including optimization of treatment. The objectives of this project are therefore broken down into 2 main axes to which will be added an axis dedicated to our objectives

in terms of dissemination and an axis dedicated to the coordination of the project. A Gantt chart of the project is provided in Section 2.2.

**Axis 1: "Collective" approach to care trajectories (Details relegated to Section 2.1.3)**

- Study the relevance of dynamic Bayesian network approach as a model for healthcare pathway and his performance for analysis,
- Development of tools to assess and control the stability of groups in the context of classification of healthcare trajectories,
- Methodological recommendations for studying the transition from a "short-term" approach to a "whole life" approach, by means of healthcare pathways simulation of Agent Based Model type,
- Make recommendations on relevant data sources and on the sizing of such studies,
- Application on medico-economic data.

**Axis 2: "Individual" approach to care trajectories (Details relegated to Section 2.1.3)**

- Study the relevance of avenues for improving machine learning techniques in the context of "reinforcement learning" for the analysis of healthcare pathways,
- Issue methodological recommendations for the use of these techniques,
- Make recommendations on relevant data sources and on the sizing of this type of study,
- Applications on clinical data.

**Axis 3: Dissemination (details relegated to Section 2.1.4)**

- Development of R package(s) specific to healthcare pathways,
- Rshiny application development of optimal processing algorithms,
- Participation in conferences and workshops,
- Organization of an end-of-project workshop.

**Axis 4: Project coordination (details relegated to Section 2.1.4)**

### 2.1.2 Positioning of the work in the context of current knowledge

**Axis 1: "Collective" approach to healthcare pathways: classification of healthcare pathways**

The classification of healthcare pathways is usually broken down into two stages: a step consisting of a measure of dissimilarity between the trajectories and a classification method based on this measure. The standard approach consists in performing optimal matching coupled with a cluster analysis. Optimal matching is a method for the analysis of sequential data to assess the dissimilarity of ordered arrays that usually represent a time-ordered sequence. In particular, it allows identifying patterns in healthcare pathways. Once such distances have been calculated for a set of observations, classical tools such as ascending hierarchical classification are used to derive a clustering of the trajectories. This approach is relatively well known and has given relevant results [1]. However, methodological obstacles are still relevant especially the performance of dissimilarity measures are relative and strongly constrain the definition of the trajectory, the classification algorithms could benefit from innovations in machine learning and the question of the stability of the groups built is often little or not measured nor discussed [2,3].

In sequence analysis, hidden Markov models (HMM) have proven to be a highly effective way of modeling a family of unaligned sequences in particular in Biology [4]. Indeed, probabilities (or costs) are associated with each character omission and each transition between states. The alignment of a sequence is simply the highest-probability (or lowest-cost) path through the HMM. In the meantime, BNs (Bayesian Networks) which allow to represent the dependency structure of a set of variables via a directed acyclic graph are ideal to study complex interaction between variables. These models are used in a wide range of applications [5] but are not suited to deal with temporal dependencies. A dynamic Bayesian network (DBN) extends the notion of Bayesian network by making it possible to represent the evolution of random variables as a function of a discrete sequence, for example time steps [6]. This tool then allows to model sequences of variables. In fact they can model the relationships between multiple time series in the same model. It was shown that DBN generalize several probabilistic models and especially the HMM. It is then natural and tempting to explore the use of BNs and DBNs to study healthcare pathways.

The question of the analysis horizon, that is to say the observation window of the healthcare trajectory is a very important question with major repercussions on the volume of data to be collected, on the performance of the algorithms and ultimately the interpretability of the results obtained. The most relevant objectives would be

expressed in long time (whole life) but usually studies are done over short horizons (a few years) for lack of data on long time scales. Little work has been initiated in this spirit, let us nevertheless cite [7] who developed an automated radiation adaptation protocols for NSCLC (Non–Small-Cell Lung Cancer) patients by using GAN (generative adversarial network) to generate synthetic patient data and DQN (Deep Q-Networks) to learn dose decisions with the synthesized data and the available real clinical data.

**Axis 2: "Individual" approach to healthcare pathways: precision medicine**

In medical research, attention focuses on dynamic treatment regime (DTR) (AKA adaptive intervention, adaptive treatment strategy). DTR is a set of rules for choosing effective treatments for individual patients based on that individual's characteristics and history. The goal consists in finding an optimal decision rule in terms of long-term clinical outcome. Since the pioneer works, (see [8,9] and references therein), this field of research is very active essentially on the pulse of the activity of one of the most popular underlying statistical frameworks: reinforcement learning (RL). Reinforcement learning consists of learning, from data, the actions to be taken so that the reward acquired over time is maximum. We consider that the agent is immersed in an environment, and makes his decisions according to his current state. In return, the environment provides the agent with a reward (positive or negative). The agent seeks to learn a decisional behavior (called policy), a function associating with the current state the action to be executed in an optimal way, in the sense that the sum of the rewards over time is maximum.

"Model-free" and "Model-based" is one of the most important branching points in an RL algorithm. "Model-based" approach consists in the setting where the agent has access to (or learns) a model of the environment, this means, a predictive model for state transitions and rewards. "Model-free" approach does not necessitate such a model, the strategy to get the optimal policy is purely data-driven. This is an essential distinction because the "model-free" approach has the undeniable advantage of being free from the estimation of the underlying Markov Decision Process and the inherent assumptions. In return, the weakness is the propensity of these strategies to explore the realm of possibilities, especially in a context of a continuous state / action space. This is known as the exploration-exploitation dilemma. This dilemma is a fundamental problem in reinforcement learning. Agent is frequently faced when choosing between options, rather:

- pick something familiar in order to maximize the chance of getting what you wanted,
- pick something not tried and possibly learning more, which may (or may not) result in making better decisions in future.

This trade-off will affect either agent earn his reward sooner or he learns about the environment first then earn his rewards later. Most of the RL approaches in health are "model-free" approaches based on Q-learning techniques. These are computationally intensive techniques and, to be relevant, require a sufficient volume of data requiring to be constrained to simple topologies for the spaces of states and actions (limited number of states, discretized).

Now, phenomenal breakthroughs in machine learning point to the applicability of advanced RL techniques in the DTR context [10]. First, the question of the extension to more complex situations in terms of state / actions spaces requires the use of learning techniques which are more complex. At the center of these learning techniques, deep Q-learning algorithms [11] are promising and will be explored in this project to evaluate theirs performances in the framework of healthcare pathways. Second, in recent years, several more advanced exploration strategies have been proposed to deal with the question of the exploration-exploitation dilemma. It is in particular possible to introduce expert a priori or at least hypotheses in the learning methods in order to restrict the exploration space and to force the learning to be limited to the most probable area. These methods are part of the PAC (probably Approximately Correct) methods [12] guaranteeing a targeted exploration of the learning algorithm switching from purely data-driven paradigm to a hybrid data / expert driven one. The contribution of these methods in the context of care pathways will be explored within the framework of this project. Third, one of the main limitations in the application of these methods is, on the one hand, the volume of the data and, on the other hand, the computational constraints. Careful attention to small sample techniques is imperative. While still in its early stage, significant progress has been made in small sample learning research in recent years [13]. As such, a direct solution is the use of "data augmentation" strategies such as deformations [14] or GANs (Generative Adversarial Network [15]) to increase samples and then employ conventional learning methods. Finally, it should be noted that the constraints linked to the formulation of the reward. Usually, the reward is defined by discretizing quantitative variables in an ad-hoc manner by the clinician. In addition, rewards data is often sparse due to the lack of evaluation at the time of decisions. These weaknesses are not or are little considered in the usual methods.

### 2.1.3 Detailed description of the methodology and of the techniques implemented

**Axis 1: Detailed description of investigation on the "Collective" approach.**

A popular approach for dealing with healthcare pathways consists in measuring the dissimilarities between the trajectories (often using optimal matching) and then to find a clustering (often using an ascendant hierarchical

classification) [1]. Although optimal matching remains the most used dissimilarity measure, there exists many other ways of measuring dissimilarity. The performance of dissimilarity measures is relative and it appears that the choice of a measure depends on which aspect we want to focus on (difference in sequencing, timing, duration, …). Choosing a suitable measure for given research objectives is a difficult task which need to be studied. In particular, the behavior of dissimilarity measures should be evaluated in different contexts, for instance, when one have to cope with sequences of unequal length or with missing elements in the sequences. Regarding clustering algorithms, innovations in machine learning could benefit to this approach by improving the clustering of sequences. For instance, methods based on fuzzy clustering could be interesting since it is a form of clustering in which each data point can belong to more than one cluster. Finally, the question of the stability of the groups built is often little or not measured nor discussed [2,3]. However, the solutions obtained are often different according to the choice of dissimilarity measures or clustering methods. The stability could be evaluated by using different dissimilarities measures, different clustering methods but also by resampling using several schemes (bootstrap, sub-setting, jittering and replacement of points by noise). One can then study the Jaccard dissimilarity or Rand index to compare original clustering with most similar clusters in resampled data in order to derive an index of the stability of a cluster [16]. The issue of stability could also be evaluated by studying variations of markers or principal components derived from preliminary factorial analysis. For each observation one could also derived indicators based on the probabilities of affectation in each cluster. The development of stability indicators could be of great interest from a practical point of view to evaluate the stability of the solution provided by a cluster analysis.

Bayesian networks (BN) are a type of probabilistic graphical model that can be used to build models from data and/or expert opinion. BNs are graphical representations of a joint probability distribution that take the form of a network made up of nodes and edges representing variables and the influences between them [5]. They are directed acyclic graphs whose nodes represent variables in a Bayesian context: they may be observable quantities, latent variables, unknown parameters or hypotheses. Edges represent conditional dependencies whereas nodes that are not connected represent variables that are conditionally independent of each other. The joint probability distribution factorizes into conditional probability distributions associated with each node conditional on variables that directly influence it. Structural learning algorithms are involved to determine the optimal factorization (based on data and/or expert opinion) which corresponds to a representation of the dependencies structure between variables. Once the structure has been defined, inference step consist in computing the probability of each state when others variables are known. Although BNs have very good risk prediction performance, their main advantages are their intuitiveness and their ability to represent the joint probability distribution as a network structure. They allow for ease of individual-level risk prediction via “what-if” scenarios (“diagnostic” reasoning) since inference is performed not only from cause to effect, as in standard predictive models (inferring from association between a disease and a symptom), but also from effect to cause (calculating the probability of having disease after having observed a symptom). Moreover BNs can incorporate individual-level or aggregate data, expert opinion or evidence from literature and can be updated with new information and maintain a high degree of flexibility to accommodate developments in knowledge, new interventions, and database size and complexity. They are then very flexible tools for “precision medicine” demanding individual risk prediction (Axis 2). Finally, BNs also have the ability to incorporate decision and utility nodes in the model (Decision graph) [17]. The objective is to find an optimal strategy (set of decisions) which maximizes the overall utility (e.g. profit/loss) of the network. This allows conducting cost-effectiveness or cost-utility analysis based on models learned from data and/or built from expert opinion.

Healthcare pathways data are sequential by nature and BN seems not adapted since they are not design to deal with the temporal dependencies. Dynamic Bayesian networks (DBN) extend the BN by allowing the evolution of random variables in function of a sequence of discrete times [6]. DBN deal with time but the model linking the different variables does not necessarily change dynamically. Indeed, a DBN consists of a series of time slices that represent the state of all the variables at a certain time, t. For each temporal slice, a dependency structure between the variables at that time is defined. Additionally, there are edges between variables from different time slices. In medical diagnosis, an example of a Bayesian network would be to estimate the risk for a patient having a disease based on his symptoms. This network can then be made “dynamic” by integrating the fact that the probability of being ill at time t also depends on the past probability. Intuitively, this means that the risk evolves over time. In a DBN, several tasks should be investigated. The inference step which consists in estimating all the unknown states in the network as only a subset of states can be observed at each time slice. The learning step consists in estimating parameters of a DBN, given several sequences of observations, such that they best fit to the observed data. Finally, the pruning step aim to identify which nodes is important for inference and to remove less important nodes from the network. This is a complex task which is often ls investigated.
Since DBNs generalize several probabilistic models applied in sequence analysis like Markov chains models or HMMs, they seem to be flexible and efficient tools to deal with different kind of healthcare pathway. For instance, these models can be used for predicting future observations or hidden states in the next time slice based on the past data or to find the most likely sequence of hidden variables given the observations. The ease to derive

individual-level risk prediction together with their capacity to deal with complex dependencies structures make DBN a promising tools in a precision medicine context (Axis 2). From a "collective" point of view, DBNs can be used after grouping sequence using sequence analysis and a clustering method. A DBN can then be fitted for each cluster for compressing information by finding time-varying latent structures (identify typical health stages within each group, pattern recognition), for comparing different clustering solutions and for performing cost analysis. Such approach is sensitive to the original clustering. Another possibility could be to consider mixture of HMMs (or DBNs) in order to fit one model to the whole data and determine clustering during modeling [18]. Instead of fixing individuals to the clusters defined during the sequence analysis, a mixture of HMMs use all data to estimate the model considering each individual belongs to each cluster with some probability.

Agent model simulation approaches [19] would make it possible to simulate long-time healthcare trajectories by exploiting the results of short-time trajectory analyzes. Such an approach makes possible to compute credibility intervals of parameters of interest deduced from simulated posteriori distribution. This strategy was used by Team 1 and 2 to estimate the differential costs generated by the switch to generic antiretroviral treatments [20]. In this context, the ABM may split in two main parts: the virtual patient's generator and the execution models [21]. The virtual patient's generator [22] aims to simulate vectors of covariates defining "virtual patients" and executions models aim to mimick patient's life involving especially healthcare pathways modeling obtained by optimal matching, by Markovian models, semi-Markovian models or DBNs introduced in Axis 1. These models are calibrated on short-time data but in this context of ABM it is easy to extend the horizon of analysis and to get an idea of what would happen in the long term. Other execution models may be involved for instance covariates' evolutions in time, comorbidities, …

As an application of these investigations to real datasets, we will implement optimal matching methods on two different medico-economic studies. First, we will use the OM methods to characterize the healthcare pathway and costs over a 5 years period of incident cases of patient with brain injuries. We will use the French Health National Data System (SNDS: Système National des Données de Santé) to record healthcare consumption of 35 000 patients managed at the university Hospital of Toulouse for brain injuries. This project is currently in progress and funded by the Health Data Hub. Second, we will use data from the SNDS to study the health care pathways and costs of patients with Chronic Myeloid Leukemia (CML). In this project, our aim is to use a national extraction of CML incident cases from the French national health insurance databases (called the "Système National des Données de Santé") to study care pathways of CML patients from 2012 to 2020, to identify determinants explaining these pathways and to evaluate the impact of the Covid-19 lockdown on the care continuum and costs. A better understanding of care pathways in CML management and the impact of the lockdown in France on its management will be crucial to prioritize future actions in public health in order to optimize the management of this costly disease.

### Axis 2: Detailed description of investigation on the "Individual" approach.

The DTR approaches [8,9] mentioned in the previous paragraph are based on reinforcement learning techniques. Basic reinforcement learning is modeled by a Markov decision process (MDP) involving

- **S** a set of environment and agent states,
- **A** a set of actions,
- $\boldsymbol{P_a}(s,s') = P(s_{t+1} = s' \mid s_t = s\,, a_t = a)$ the probability of transition (at time *t*) from state *s* to state *s'* under action *a*,
- $\boldsymbol{R_a}(s,s')$ the immediate reward after transition from state *s* to state *s'* under action *a*.

Agent interacts with its environment in discrete time steps. At a given time *t,* his current state $s_t$ and his reward $r_t$ are collected. The agent chooses an action $a_t$ and sends it to the environment which moves to state $s_{t+1}$ and reward $r_{t+1}$ according to the transition $\boldsymbol{P_a}$.
One aim on major interest is to learn a so-called optimal policy which is the map $\pi*: \boldsymbol{A} \times \boldsymbol{S} \to [0,1]$ where $\pi*(a,s) = P(a_t = a \mid s_t = s)$ and which maximizes the expected cumulated reward $\pi*(a,s) = \underset{\pi}{\operatorname{argmax}}\, Q^{\pi}(a,s)$ where $Q^{\pi}(a,s) = E_{\pi}[R_t \mid a_t = a, s_t = s]$. Q-learning is the most popular model-free, off-policy algorithm to find optimal policy and is based on Bellman Equation. Usually applied to the context of discrete state / action spaces, the optimization procedure is based on Q-table as illustrated in Figure 1 below. In more elaborated setting, especially involving covariates, Q-table is replaced by a approximating Q-function fitted on the dataset.

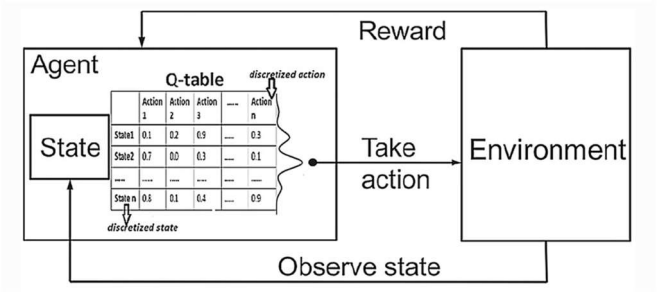


*Figure 1: Q-learning [23].*

In the medical context of long-term patient care, the reinforcement learning setting can be described as follows. For each patient, the stages correspond to clinical decision points in the course of the patient's treatment. At these decision points, actions (e.g., treatments) are chosen, and the state of the patient is recorded. As a consequence of a patient's treatment, the patient receives a (random) numerical reward concretized by a numerical outcome (survival times for instance).

Three mains weakness are currently observed in Reinforcement Learning approach and will be investigated in this project: How to approach more complex situations in terms of space of states and action? How to quantify the exploration-exploitation dilemma? How to approach the issue of small databases?

First, the question of extension to more complex situations requires the use of learning techniques that are themselves more complex. As part of this project to explore the performance and feasibility of alternative methods at the heart of which deep Q-learning (DQL) [11]. In deep Q-learning, we use a neural network to approximate the Q-value function. The state is given as input and the Q-value of all possible actions is generated as output. Figure 2 illustrates the algorithm and can be compared to Figure 1 to identify the contribution of Deep Learning. Thanks to this modeling, deep Q-learning allows to explore much more complicated setting as Q-learninq but like most deep learning algorithms, deep Q-learning presents many technical subtleties that we plan to explore within the framework of this project. The aim is to propose a clear methodology on the potential uses of this technique and its ramifications [24] in the context of the healthcare pathways analysis.

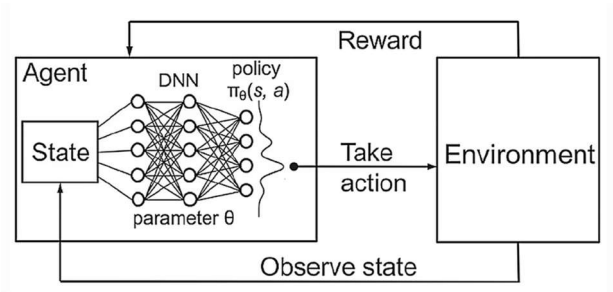


*Figure 2: Deep Q-learning [23].*

Second, there is also the question of exploring the space of possible states and actions. This problem arises at two levels: at the level of learning where only the domain defined by the data is used for learning and at the level of prediction, the optimal trajectory not necessarily being observed in the data and we use therefore a proxy of this trajectory called "off-policy evaluation". This is all the more a problem in terms of health, the optimal trajectory may be unobservable (cost, risk of uncontrolled treatment, ….). There is now ample evidence that without explicit tailoring, learning can lead to biased, unsafe, and prejudiced solutions. The question of policy evaluation from observational data is therefore an important methodological issue to be discussed. To address this problem, we plan to use techniques allowing to introduce experts' a priori into learning techniques moving from a purely data-driven paradigm to a hybrid one mixing data-driven and expert-driven techniques. One method attracts our attention in this context and will be studied within the framework of this project is the PAC (Probably Approximately Correct) method [12]. In the context of PAC learning, the "learner" algorithm receives training data ("samples") and must choose a function which generalizes these data. This function is chosen from a pre-established set possibly by experts' a priori. The goal is that the function in question classifies new unknown data (distributed identically to the training data) with minimal error, and this with high probability. Finally, the question will benefit from advances in PAC techniques, in particular Probably Approximately Correct Constrained Learning [25].

Third, like any data-based learning approach, the performance of these methods will be directly related to the volume of data. In the context of healthcare pathways, to ensure that we have data of quality, it is tempting to sacrifice massive databases and limit ourselves to small databases. To ensure the applicability of the techniques developed above, it is nevertheless necessary to think about strategies to overcome these constraints of dimension. To do this, a direct approach is the use of "data augmentation" strategies. The idea is relatively naive

and consists in artificially increasing the volume of the database by generating synthetic patient's data, then conventional learning techniques remain to be applied. Methods of virtual patient generations have already been developed by team 1 [21,22], in particular using R-vine-type copula methods. As part of this project we plan to enrich virtual patient models using GANs (Generative Adversarial Network). A GAN is a machine learning technique based on the competition of two networks called "generator" and "discriminator". The generator is a type of convolutional neural network whose role is to create new instances of an object (a virtual patient in our setting). On the other hand, the discriminator is a "deconvolutive" neural network that determines the authenticity of the object or whether or not it is part of a data set. During the training process, these two entities compete and this is what allows them to improve their respective behaviors. This is called backpropagation. The objective of the generator is to produce outputs without being able to determine whether they are false, while the objective of the discriminator is to identify false ones. Thus, over the course of the process, the generator produces better quality outputs while the discriminator detects the false ones better and better. In fact, the illusion becomes more and more convincing over time. Initially used for images generation, GANs are more and more popular and is of capital interest for virtual patients' generation. The main questions that we will investigate in this project is, first, to study the quality of the data generated by GANs especially in terms of structural dependence between covariates and, second, to study and compare these "data augmentation" approaches to understand the analysis of care pathways from small samples.

Finally, the counterpart of these investigations is the introduction of more and more sophisticated learning techniques and therefore more and more opaque black boxes. Particularly in the context of health data, it is necessary to attach particular importance to the interpretability of the results. Avenues are open [26,27] and will be explored within the framework of this project.

These investigations will be applied to real dataset. The context of interest in this project is chronic lymphoid leukemia. In this context, DTR has been initially performed in [28] by studying a two-stage clinical decision: initially prescribe the first-line induction therapy a; if the patient responds to a, prescribe maintenance therapy, and if the patient does not respond to a, prescribe the second-line induction therapy b. In practice things are much more complicated and necessitate much more complicated models as those investigated in this project. Indeed, in chronic lymphoid leukemia, first line treatment mainly involves immune-chemotherapy or targeted agents, depending on both the patients' mutational profile, age, and comorbidity, in addition with prophylactic treatment. In second line, kinase inhibitors or anti-BLC2 antibody are possible. They both have different safety and efficacy profiles, rendering critical the medical decision for any individual patient. Response to treatment is assessed on physical examination and evaluation of the blood and bone marrow, 3 months after the last administration of immunochemotherapy or during continuous targeted therapy. The choice of therapeutic strategy is bounded by clinical guidelines, which nevertheless leave some decision latitude between equivalent alternatives depending on the disease and the patient's characteristics (co-morbidity, mutation profile, etc.). In observational setting, i.e. outside clinical trials, the complexity of the interrelation between factors that may affect treatment, as well as the variety of these factors, may challenge the assessment of the optimal care trajectory. Each year, the Cancer University Institute of Toulouse follows about 1000 CLL patients among which 300 are treated. It is thus possible to build a specific cohort of patients to assess the performances of the model introduced below and to construct personalized optimal decision rules

### 2.1.4. Plan for executing the project

- **The role of each team**

**Team 1** is composed of researchers in applied mathematics with different profiles. Nicolas Savy specializes in statistical methodology applied to health, Philippe Saint-Pierre specializes in health data analysis and machine learning and Jean-Michel Loubès is a specialist in artificial intelligence. The mission of this team will be to reflect on the methodology for using the methods presented in Section 2.1.1, to implement them and to develop alternative and / or complementary methods. To do this, team 1 will receive support of two international experts who have accepted to collaborate on this project to share their skills on Dynamic Treatment Regimes by organizing "research stays" and to provide an external perspective on the relevance of the methodological and strategic choices.

- **Michael Kosorok (Chapel Hill University (USA))**. He is chair of the department of Biostatistics of Chapell Hill University. His research expertise is in empirical processes, biostatistics, data science, machine learning and precision medicine. He also has expertise in the application of biostatistics and data science to human health research, including cancer and cystic fibrosis. He has pioneered machine learning and data mining tools for these and related areas.
- **Erica Moodie (McGill University (CAN))**. She is Professor of Biostatistics at McGill University. Her research expertise is in Biostatistics, causal inference and longitudinal data analysis with a focus on adaptive treatment strategies.

**Team 2** “Health Economics” performs research into the determining factors, management options and consequences of chronic diseases and disabilities, within a constantly changing environment (financial constraints, health security, increasing inequalities). The team has a large experience in conducting economic analysis in randomized controlled trials in chronic diseases, and on cancers (see for example [29,30]). It is currently involved in different national programs such as PHRC and PRME programs. This team has expertise on the exploitation and analysis of medico-administrative database (SNDS). Finally, this team led several projects with Team 1, which guarantees the feasibility of the project [20].

**Team 3** « cancer clinical epidemiology» involves epidemiologists from the Claudius Regaud Institute and Hematologist from the Toulouse University Hospital, to provide a clinical expertise on the cancer care and patients care trajectory, more specifically in Onco-Hematology. Both are active members of the Onco-Occitanie regional cancer healthcare network, which coordinate multidisciplinary team meetings (MTM) dedicated for therapeutic decision making regarding all hematological malignancies of the 3 million inhabitants Ouest-Occitanie area. Moreover, this teams will ensure the building, management, and preparation of the clinical database dedicated to the analyses of the individual-level care trajectory. This team have a strong experience in studying the cancer patients’ care trajectory, including in onco-hematology, as well as the factors influencing the clinical decisions.

- **The identification of the key stages**

The key stages are detailed below and the timeline is recorded in the Gantt chart (Figure 3).

**The key stages of axis 1:**

- Step 1: DBN approach to healthcare pathways analyses
    a) Investigations on the relevancy of DBN approach in the context of healthcare pathways analyses
    b) Performance assessment and comparison based on data on costs associated with thyroid cancer
- Step 2: Assessment of the stability of the groups built by the Optimal Matching / CAH approach
    a) Bibliographical synthesis of existing techniques and identification of relevant techniques
    b) Performance assessment and comparison based on data on costs associated with thyroid cancer
    c) Implementation in the R package
- Step 3: Investigation of the use of agent models (ABM) for the transition from short to long horizon.
    a) Collection of clinical constraints / Identification of the execution models to build and of the virtual patient generator
    b) Identification and collection of data for the configuration of the ABM
    c) Implementation of ABM for costs associated with thyroid cancer / Definition of the scenarios / Interpretation of results, visualization of results
    d) Construction of an R package on ABMs for the treatment path
- Step 4: Study of a new "case study": brain injuries and LMC.
    a) Collection of clinical constraints - Identification of execution models to be built and of the virtual patient generator - Reflection on the scenarios to be investigated
    b) Identification and collection of data for the configuration of the ABM
    c) Use of the package for obtaining results and visualization

**The key stages of axis 2:**

- Step 1: Deep Q-learning to explore more complicated decision rules.
    a) Investigation of Deep Q-learning models (This step will be carried out upstream of the project via an M2 internship in 2021)
    b) Research of the format, the nature, the volume of the relevant data for the implementation of DQN
    c) Application of DQN on real dataset, interpretation and feedback
- Step 2: Hybrid data/expert driven approach for healthcare pathways analyses
    a) Recommendations on the suitability of the model / objective / data
    b) Application to real data: construction of optimal decision rules
    c) If the rules are relevant, building an Rshiny application for clinicians
- Step 3: “Individual” approach from small databases
    a) Use of GANs and GPVs for the "data augmentation" approach from small databases
    b) Identification of the thresholds of relevance of the algorithms according to the base and identification of the discriminating performance factors
- Step 4: Construction of a dedicated databases for implementation of those approaches on real-world data

- **The methods for coordinating the project and for quality control**

The teams of this consortium have a long experience of working together on several projects. The PIs of each team will be in charge of information within each of the teams, especially for Team 3, Sébastien Lamy will follow up on discussions with the clinicians involved in the project. The "day-to-day" monitoring of the project will be ensured by Romain Demeulemeester, currently a doctoral student in team 2 in joint supervision between Teams 1 and 2, and who would work as researcher for this project.

Quarterly summary meetings will be organized to present the results in progress, to decide collectively on the direction of the project, on the content of the publications of the results obtained. At the end of the project, these meetings will be more spaced out to make room for organizational meetings for the end of project workshop. To help us in the conduct of this project, our two international experts will be invited to participate by videoconference in our quarterly meetings.

### 2.1.5 Expected results, as well as possible spin-offs for them.

The target audience of ML4PHA project is the whole Health community whose purpose is to understand or optimize healthcare pathways. This audience is not necessarily informed of the methodological advances, especially the very innovative ones stemming from the statistics. Moreover, this public is not necessarily seasoned with the software techniques and asks that efforts be made to propose easy-to-use, user-friendly and interactive packages (Rshiny for example).

Regarding the communication of the results, it will be divided into:

- Publications in statistical journals but also in the most general journals possible to sensitize the most varied communities. The publications will be systematic deposite on HAL.
- Publications in specialized journals of health economical results from investigations of Axis 1.
- Publications in specialized journals of clinical results from investigations of Axis 2.
- Participations to scientific conferences (statistics, biostatistics and medical) especially JSM conferences in 2022 and 2024.
- Building of Rshiny apps to help clinicians to use the results from investigations of Axis 2.
- The dissemination of the proposed algorithms through R packages with a special effort on the user-friendly side of the application with the deployment of tools like RShiny.
- The setting up of a web page dedicated to the project and presenting in a vulgarized way the possibilities offered by these algorithms.
- The organization of a closing workshop.

**References**


[1] G. Ritschard and M. Studer (Eds.). *Sequence Analysis and Related Approaches Innovative Methods and Applications.* Life Course Research and Social Policies, SpringerOpen editors, 2018.

[2] M. Studer and G. Ritschard. *What matters in differences between life trajectories: a comparative review of sequence dissimilarity measures?* J. R. Statist. Soc. A 179, Part 2, pp. 481–511, 2016.

[3] R. Piccarreta, M. Studer. *Holistic analysis of the life course: Methodological challenges and new perspectives*. Advances in Life Course Research, Volume 41, 2019.

[4] R. Hughey and A. Krogh. *Hidden Markov models for sequence analysis: extension and analysis of the basic method.* Bioinformatics 12(2), 95-107, 1996.

[5] P. Weber, G. Medina-Oliva, C. Simon, and B. Lung. *Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas*. Engineering Applications of Artificial Intelligence, 25(4), 671-682, 2012.

[6] M.T. Amin, F. Khan and S. Imtiaz. *Fault detection and pathway analysis using a dynamic Bayesian network*. Chemical Engineering Science, 195, 777-790, 2019.

[7] H.H. Tseng, Y. Luo, S. Cui, J. T. Chien, R.K. Ten Haken, and I.E. Naqa, *Deep reinforcement learning for automated radiation adaptation in lung cancer*. Medical Physics, vol. 44, no. 12, pp. 6690–6705, 2017.

[8] M.R. Kosorok and E.E.M. Moodie (Eds). *Adaptive Treatment Strategies in Practice*. ASA-SIAM Series on Statistics and Applied Mathematics, 2015.

[9] B. Chakraborty and E.E.M. Moodie. *Statistical Methods for Dynamic Treatment Regimes: Reinforcement Learning, Causal Inference, and Personalized Medicine*. Springer. 2013.

[10] Y. Chao, L. Jiming, and N. Shamim. *Reinforcement learning in healthcare: A survey*. arXiv preprint arXiv:1908.08796, 2019.

[11] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare and J. Pineau, *An Introduction to Deep Reinforcement Learning*. Foundations and Trends in Machine Learning: Vol. 11, No. 3-4, 2018

[12] J. Pazis and R. Parr, *PAC optimal exploration in continuous space Markov decision processes*. in AAAI, 2013.

[13] S. W. Carden and J. Livsey, *Small-sample reinforcement learning: Improving policies using synthetic data*. Intelligent Decision Technologies, vol. 11, no. 2, pp. 167–175, 2017.

[14] J. Salamon and J. P. Bello, *Deep convolutional neural networks and data augmentation for environmental sound classification*. IEEE Signal Processing Letters, vol. 24, no. 3, pp. 279–283, 2017.

[15] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial nets*. Advances in neural information processing systems, pp. 2672–2680, 2014.

[16] C. Hennig. *Cluster-wise assessment of cluster stability*. Computational Statistics and Data Analysis, 52, 258-271, 2007.

[17] P. Arora et al. *Bayesian Networks for Risk Prediction Using Real-World Data: A Tool for Precision Medicine*, Value in Health, 22(4), 439-445, 2019.

[18] R. Altman. *Mixed Hidden Markov Models: An Extension of the Hidden Markov Model to the Longitudinal Data Setting*. JASA, 102(477), 201-210, 2007.

[19] S. Railsback and V. Grimm, *Agent-based and individual-based modeling: a practical introduction*. Princeton university press. 2019.

[20] R. Demeulemeester, N. Savy, M. Mounié, L. Molinier, C. Delpierre, P. Dellamonica, C. Allavena, P. Pugliese, L. Cuzin, P. Saint-Pierre, N. Costa. *Economic impact of generic antiretrovirals in France for HIV patients' care: a simulation between 2019 and 2023.* in submission in value in health, 2021.

[21] N. Savy, P. Saint-Pierre, S. Savy, S. Julien, E. Pham, *" In Silico Clinical Trials": a way to improve drug development?*. Proceedings of JSM 2019 – Biopharmaceutical Section, Denver, 2019.

[22] N. Savy, S. Savy, S. Andrieu, S. Marque. *Simulated Clinical Trials: Principle, Good Practices, and focus on Virtual Patients Generation.* Springer Proceedings in Mathematics & Statistics, chapter 21, 2018.

[23] R. Jafari and M. Javidi. *Solving the protein folding problem in hydrophobic-polar model using deep reinforcement learning.* SN Appl. Sci. 2, 259, 2020.

[24] K. Arulkumaran, M.P. Deisenroth, M. Brundage, A.A. Bharath. *A Brief Survey of Deep Reinforcement Learning.* IEEE signal processing magazine, 2017.

[25] L.F.O. Chamon and A. Ribeiro. *Probably Approximately Correct Constrained Learning*. arXiv:2006.05487v1, 2020.

[26] Z. Che, S. Purushotham, R. Khemani, and Y. Liu. *Distilling knowledge from deep networks with applications to healthcare domain.* arXiv preprint arXiv:1512.03542, 2015.

[27] M. Wu, M.C. Hughes, S. Parbhoo, M. Zazzi, V. Roth, and F. Doshi-Velez, *Beyond sparsity: Tree regularization of deep models for interpretability*. Thirty-Second AAAI Conference on Artificial Intelligence, 2018.

[28] A.S. Wahed, and A.A. Tsiatis. *Optimal Estimator for the Survival Distribution and Related Quantities for Treatment Policies in Two-Stage Randomization Designs in Clinical Trials*. Biometrics, 60: 124-133, 2004.

[29] M. Mounié, N. Costa, C. Conte, D. Petiot, D. Fabre, F. Despas, M. Lapeyre-Mestre, G. Laurent, N. Savy and L. Molinier, *Real-world costs of illness of Hodgkin and the main B-Cell Non-Hodgkin lymphomas in France*. Journal of Medical Economics, 23:3, 235-242, 2020.

[30] C. Canivet, N. Costa, F. Ory-Magne, C. Arcari, C. Mohara, L. Pourcel, H. Derumeaux, E. Bérard, R. Bourrel, L. Molinier, C. Brefel-Courbon. *Clinical Impact and Cost-Effectiveness of an Education Program for PD Patients: A Randomized Controlled Trial*. PLoS One. 2016.

## 2-2 Calendar for key stages

For the sake of simplicity and clarity of the project, the schedule of the key stages of the project is given by means of the Gantt chart Figure 3 below. Steps refer to the steps defined in section 2.1.4. The numbers in each box refer to the teams involved in that step.

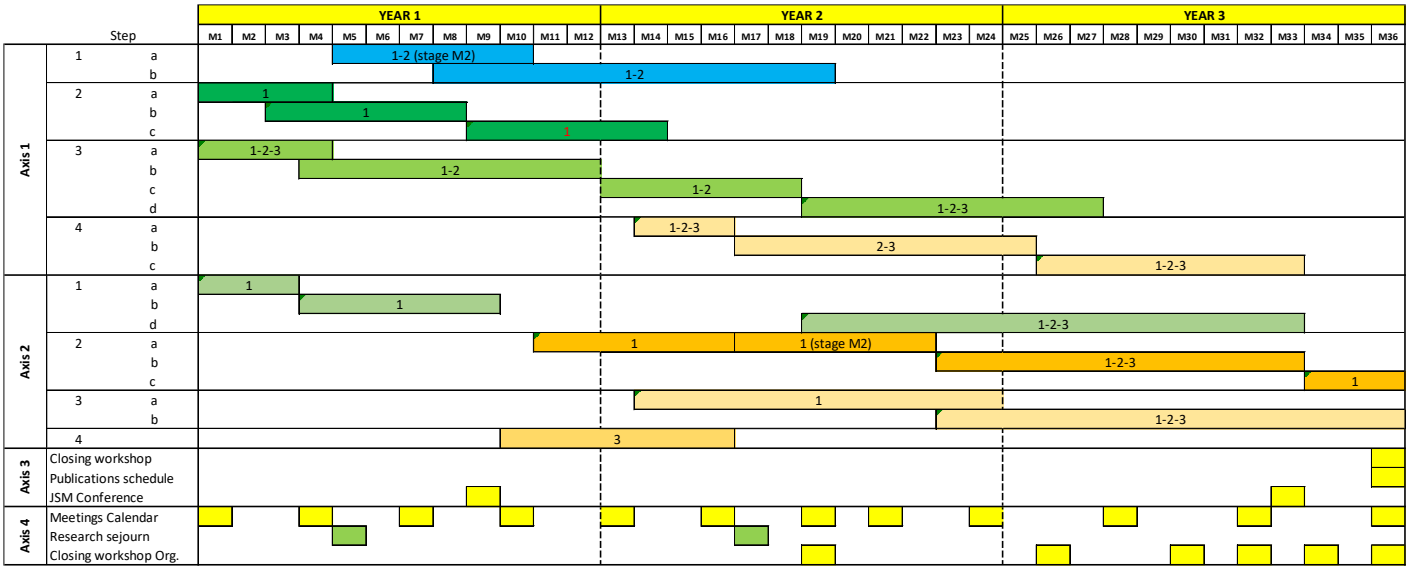


*Figure 3: Gantt chart of the ML4HPA project.*

## 2-3 Budget justification

Table below summarizes the requested budget for ML4HPA project. Justifications of main budget lines referenced in column "Justif." are given in a precise way under the Table.

| | Team | 2021 | 2022 | 2023 | 2024 | Justif. |
|---|---|---|---|---|---|---|
| Post Doc (Romain Demeulemeester) | 1 | 3 766 | 45 192 | 18 830 | - | (1) |
| Post Doc (Romain Demeulemeester) | 2 | | - | 18 076 | 18 077 | (1) |
| Clinical research assistant (CRA to hire) | 3 | | 13 836 | 18 448 | | (2) |
| Senior epidemiologist (Sébastien Lamy) | 3 | | 3 350 | 3 350 | | (3) |
| Miscellaneous travel expenses | 1 | | 1 000 | 1 000 | 1 000 | |
| Miscellaneous travel expenses | 2 | | | 1 500 | 1 500 | |
| Training stay | 2 | | | 2 000 | 2 000 | (4) |
| Participation in JSM conferences | 1 | | 3 000 | | 3 000 | (5) |
| Training stay | 1 | | 4 000 | 4 000 | | (6) |
| Organization of a closing workshop | 1 | | | | 5 000 | (7) |
| Equipment | 1 | | | 3 500 | | (8) |
| Master internship gratuity | 1 | | 3 600 | 3 600 | | |
| Operating costs | 1 | | | 1 000 | 1 000 | (9) |
| Operating costs | 2 | | 1 500 | 1 500 | 1 500 | (9) |
| Operating costs | 3 | | 2 500 | 2 500 | | (9) |
| Management fees | 1 | 302 | 4 543 | 2 554 | 800 | |
| Management fees | 2 | | 1 726 | 1 726 | | |
| Management fees | 3 | | 1 750 | 1 750 | | |
| **Total** | | **4 068** | **82 771** | **83 834** | **34 103** | **204 776** |

(1) Most of the budget is dedicated to the financing of Romain Demeulemeester. Mathematician by training, Romain is currently doing a doctoral thesis, supervised by Nicolas Savy, Pascale Grosclaude and Nadège Costa on issues of healthcare pathways (optimal matching) to assess the costs associated with thyroid cancer. In addition, Romain did his master's internship, supervised by Nicolas Savy, Philippe Saint-Pierre and Nadège Costa, on dynamic treatment regimes. Finally, Romain did a research stay in 2020 at McGill university supervised by Erica Moodie. Romain is thus a central part of this project and has all the skills to carry out this ambitious project.

(2) A CRA (to hire) will ensure the collection and construction of the dedicated database for the real-world implementation of advanced "Reinforcement Learning" tools for the analysis of healthcare pathways.

(3) This step will be done under the supervision of clinicians and epidemiologists from team 3, in particular Sébastien Lamy, part of the time of which will be dedicated.

(4) During the winter of 2023, a research stay is planned for Romain Demeulemeester at McGill University. During this stay, Romain will benefit from the skills of Erica Moodie to finalize various aspects of the project, in particular the technical aspects related to axis 2. It will also be the opportunity to work with Elisabeth Krakow, epidemiologist at McGill and specialist in leukemia.

(5) Participating in JSM conferences is a key point in the scientific communication of the project. Indeed, it is the largest conference in the world in statistics and biostatistics. It is held annually in the United States and gives rise to widely distributed proceedings. Nicolas Savy plans to participate in the 2022 editions in Washington DC and 2024 in Portland.

(6) In the context of this project, two research stays for Nicolas Savy and Philippe Saint-Pierre are planned: in 2022, a stay at Chapel Hill University with Prof. Kosorok and in 2023, a stay at McGill University with Prof. Moodie. These stays are of capital importance for the project, they allow us to benefit from the skills of the best teams from around the world and to challenge our research with them.

(7) This envelope allows the organization of a 2-day fees free workshop (lunch, dinner, coffee breaks). Co-financing will be sought from the university, CNRS, IMT, the Occitanie region, the GSO cancer center to arrive at a total envelope of € 10,000 allowing to invite speakers in particular our project partners Erica Moodie and Michael Kosorok.

(8) Acquisition of 2 laptops.

(9) Includes publication costs, teams' coordination costs, and database setup costs for Team 3.

### 2-4 Skills and/or expertise in the area of the call for proposals

An English version of the CVs of the coordinator and of each team manager is provided as an attached single file document.

## Part III

### 3-1 Project Coordinator

| **Signature du coordinateur de projet / *Signature of the project coordinator*** |
|---|
| Je soussigné, (nom et prénom du coordonnateur de projet) :<br>SAVY Nicolas<br>confirme la faisabilité du projet tel qu'il a été décrit dans le dossier de candidature.<br><br>**I the undersigned, (last and first name of the project coordinator):**<br>**SAVY Nicolas**<br>**confirm the feasibility of the project such as has been described in the application file.**<br><br>Signature : Fait à Toulouse, le 08/02/2021<br>Signature: Signed in Toulouse on 08/02/2021 |

| **Signature du directeur d'unité qui héberge le coordonnateur de projet / *Signature of the director of the laboratory that hosts the project coordinator*** |
|---|
| Je soussigné, (nom et prénom directeur du laboratoire) :<br>BARTHE Franck<br>autorise M ~~ou Mme~~ Nicolas SAVY<br>à développer et coordonner ce projet et m'engage à lui en permettre la réalisation.<br><br>**I the undersigned, (last and first name of the director of the laboratory):**<br>**BARTHE Franck**<br>**authorize Mr ~~or Mrs~~ Nicolas SAVY**<br>**to develop and coordinate this project and agree to allow this person to carry it out.**<br><br>Signature : Fait à Toulouse, le<br>Signature: Signed in Toulouse on |

### 3-2 Associated teams

| **Signature du responsable de l'équipe 2 / *Signature of the manager of Team 2*** |
|---|
| Je soussigné, (nom et prénom):<br><br>Confirme ma participation au projet coordonné par :<br>En tant qu'équipe participante.<br><br>**I the undersigned, (last and first name):**<br><br>**Confirm my participation in the project coordinated by:**<br>**As a participating team.**<br><br>Signature : Fait à Toulouse, le 08/02/2021<br>Signature: Signed in Toulouse on 08/02/2021 |

| **Signature du directeur d'unité qui héberge l'équipe 2 / *Signature of the director of the laboratory that hosts the Team 2*** |
|---|
| Je soussigné, (nom et prénom directeur du laboratoire) :<br><br>autorise M ou Mme<br>à participer à ce projet et m'engage à lui en permettre la réalisation.<br><br>**I the undersigned, (last and first name of the director of the laboratory):**<br><br>**authorize Mr or Mrs**<br>**to participate to this project and agree to allow this person to carry it out.**<br><br>Signature: Fait à Toulouse, le<br>Signature: Signed in Toulouse on |

**Signature du responsable de l'équipe 3 / *Signature of the manager of Team* 3**

Je soussigné, (nom et prénom):
LAMY Sébastien
Confirme ma participation au projet coordonné par : SAVY Nicolas
En tant qu'équipe participante.

**I the undersigned, (last and first name):**
**LAMY Sébastien**
**Confirm my participation in the project coordinated by: SAVY Nicolas**
**As a participating team.**

Signature :
Signature:

Fait à Toulouse, le 10/02/2021
Signed in Toulouse on 10/02/2021

**Signature du directeur d'unité qui héberge l'équipe 2 / *Signature of the director of the laboratory that hosts the Team 3***

Je soussigné, (nom et prénom directeur du laboratoire) :

autorise M ~~ou Mme~~ LAMY Sébastien
à participer à ce projet et m'engage à lui en permettre la réalisation.

**I the undersigned, (last and first name of the director of the laboratory):**

**authorize Mr ~~or Mrs~~** LAMY Sébastien
**to participate to this project and agree to allow this person to carry it out.**

Signature:
Signature:

Fait à Toulouse, le
Signed in Toulouse on