

Fiche de poste Data Scientist
Analyse de données de cancer dans le SNDS



Identification du poste

Poste :	Data Scientist
Lieu :	Institut Curie, 26 rue d'Ulm, Paris
Type et durée du contrat :	Temps plein – 24 mois

Présentation de l'Institut Curie

L'Institut Curie, 1er centre français de lutte contre le cancer, associe un centre de recherche de renommée internationale et un ensemble hospitalier de pointe qui prend en charge tous les cancers y compris les plus rares. Fondé en 1909 par Marie Curie, l'Institut Curie rassemble sur trois sites (Paris, Saint-Cloud et Orsay) 3 700 chercheurs, médecins et soignants autour de ses trois missions : Soins, Recherche & Enseignement.

Présentation du SNDS

Unique en Europe, voire au monde, le Système National des Données de Santé (SNDS) constitue une avancée considérable pour analyser et améliorer la santé de la population. Géré par la Caisse Nationale de l'Assurance Maladie (CNAM), le SNDS permet de chaîner :

- les données de l'Assurance Maladie (base SNIIRAM)
- les données des hôpitaux (base PMSI)
- les causes médicales de décès (base du CépiDC de l'Inserm)
- les données relatives au handicap (en provenance des MDPH)

La base de données qu'est le SNDS inclut donc des données de vie réelles recouvrant la quasi-totalité de la population française.

Présentation du projet COMBICANCER

En 2018, 18,1 millions de nouveaux cas de cancers et 9,6 millions de décès par cancer ont été dénombrés dans le monde. Entre 40% et 70% des patients atteints de cancer ont de plus des pathologies associées au moment du diagnostic (comorbidités). Ces comorbidités peuvent être associées à la prise de comédications chroniques pendant la prise en charge du cancer. Des preuves épidémiologiques ont rapporté une association entre certains médicaments et une diminution du risque de développer un cancer, une baisse du nombre de cas de récurrences et une amélioration notable de la survie dans certains types de cancer.

Toutefois, ces comédications peuvent interagir avec les thérapies moléculaires ciblées (TMC) qui constituent aujourd'hui l'essentiel des nouvelles thérapeutiques en oncologie. Cependant, les interactions entre les comédications utilisées en vie réelle chez les patients atteints de cancer et les TMC n'ont pas été caractérisées de manière extensive. Des travaux ont d'ores et déjà démontré le potentiel impact, bénéfique ou délétère des comédications sur la réponse au traitement et l'évolution à long terme dans le cancer du sein. Ceci justifie donc le besoin d'études complémentaires dans d'autres types de cancers (poumon, colorectal, prostate, cancers gynécologiques...). Ces comorbidités et comédications peuvent également impacter le parcours de soin des patients atteints de cancer.

Ainsi, l'objectif du projet COMBICANCER est d'étudier les associations entre comédications, comorbidités et cancer à une très large échelle, sur l'ensemble de la population française des patients

atteints de cancer à partir de 2007, à partir de données de vie réelle issues du Système National des Données de Santé (SNDS).

Présentation du poste

Construire à partir des données du SNDS les cohortes de patients ainsi que les covariables d'intérêt pour le projet COMBICANCER.

Vous serez en charge de la construction des cohortes SNDS par type de cancers : il s'agira d'identifier un ensemble de patients traités pour les différents cancers concernés sur une période donnée à partir des données du SNDS hébergées sur la plateforme de Health Data Hub. Votre mission consistera ensuite en la construction des covariables d'intérêt (âge, type de dates de traitements reçus, prise de comédications, existence de comorbidités, sous-type de cancers...).

Le poste est à pourvoir à plein temps au sein de l'unité U932 Immunité et cancer et au sein du Département de Recherche Translationnelle, dans l'équipe Réponse au Traitement et Résidu Tumoral (RT2Lab). Cette équipe porte de nombreux projets de recherche sur des cohortes cliniques et des bases de données médico-administratives (type SNDS) fournissant un accompagnement technique et une expertise méthodologique, biostatistique, et épidémiologique.

Compétences :

- Compétences en développement R, Python et SQL
- Compétences en analyse de bases de données
- Compétences en statistiques
- Esprit d'équipe, communication, ouverture d'esprit, créativité, curiosité, transdisciplinarité
- Être autonome, avoir un bon sens de l'organisation du travail et des priorités

Connaissances associées :

- Parcours patients en fonctions des pathologies
- Monde hospitalier (structures, processus) et des métiers de la santé (dont la recherche sur données)
- Capacité à lire la littérature scientifique et à contribuer à des articles scientifiques

Qualités requises :

- Capacité à communiquer avec différents types de profils (praticiens, chercheurs, statisticiens...)
- Des qualités d'autonomie, de flexibilité et de responsabilité
- Esprit d'équipe et volonté de prendre part à une aventure collective
- Sens de l'écoute, du résultat et de la qualité

Formation requise :

Bac+5/Ingénieur en Informatique, Data Science, Mathématiques appliquées, ou équivalent avec spécialisation sur les enjeux de traitement de données massives (Big Data)

Responsable :

Dr. Anne-Sophie Hamy-Petit, MCU-PH Université Paris Cité

Anne-sophie.hamy-petit@curie.fr