

## FICHE DE POSTE

### Ingénieur statistique qualité des données au CépiDc

#### Contexte

Le CépiDc est situé à l'hôpital du Kremlin-Bicêtre (Val de Marne). Il a pour missions principales de produire les données nationales de mortalité par causes, de réaliser des études sur la qualité de ces données (600 000 certificats par an) et d'accompagner les équipes de recherche en épidémiologie sur les statistiques de décès. Ces données sont essentielles au suivi de l'état de santé des populations, et donc à la programmation des politiques de santé publique.

La production des données repose dans un premier temps sur le codage médical des causes de décès présentes sur les certificats de décès que les médecins complètent avec le processus morbide ayant conduit à la mort. Le codage des causes médicales se fait en deux étapes : l'attribution d'un code à chaque maladie, traumatisme ou cause externe de décès mentionné sur le certificat, puis, l'identification de la cause initiale de décès selon des règles de décision internationales. La Classification internationale des maladies (CIM) définit les codes, les règles et les directives permettant de mener ces tâches à bien.

Parallèlement, les certificats de décès n'étant pas nominatifs, un chaînage indirect, est réalisé avec les données d'état civil reçues par l'Insee. Ce chaînage ou « synchronisation » permet d'une part de s'assurer que le CépiDc produit la statistique sur les causes médicales de tous les décès tels que déclarés à l'état civil ; d'autre part, ce chaînage, en enrichissant la base du CépiDc, permet d'analyser les facteurs socio-démographiques de variation de la mortalité par cause.

Le CépiDc contribue au développement et à l'optimisation des outils de codage automatique et de chaînage probabiliste, fondés sur des règles et/ou sur des méthodes d'apprentissage statistique. Le poste s'inscrit pleinement dans ce contexte et se situe à l'interface entre le pôle de production des données et le pôle d'expertise statistique.

#### Missions

L'ingénieur en charge des statistiques sur la qualité du codage médical et de la synchronisation aura pour mission de :

- Analyser la distribution de l'erreur et les conséquences statistiques de la mise en place d'outils de codage et de chaînage automatique, en particulier fondés sur l'apprentissage statistique,
- Prendre en compte l'incertitude, le changement (règles, version de classification) et la complexité du codage et du chaînage dans les indicateurs statistiques issus de l'apprentissage statistique,
- Participer à l'industrialisation de ces outils, à leur intégration et à leur pérennisation dans le processus de production des données.

#### Activités

- Réaliser des traitements statistiques pour identifier les facteurs prédictifs de l'erreur de codage ou de chaînage automatique,
- Produire des indicateurs d'écart au gold standard pour identifier les catégories de causes de décès et de population les plus affectées par les biais de codage et de chaînage automatique,
- Proposer, avec le pôle production, une méthodologie optimale de ciblage des certificats de décès à reprendre par codage manuel pour minimiser l'erreur,
- Proposer et mettre en œuvre des méthodes statistiques permettant d'intégrer l'incertitude du codage ou du chaînage dans les mesures d'incertitude (variance, écart-type) des indicateurs statistiques usuels utilisés en santé publique (taux standardisés, SMR),
- Mettre en œuvre l'apprentissage statistique sur des nouveaux jeux de données.

#### Spécificité(s) du poste :

- Langues de travail : français et anglais

## Profil recherché

### **Connaissances :**

- Maîtrise des outils statistiques,
- Connaissance des principes de l'apprentissage,
- Maîtrise de la gestion de jeux de données multiples et complexes.

### **Savoir-faire :**

- Maîtrise des langages R et Python,
- Maîtrise du langage SQL,
- Rédaction de documentation technique, de rapports et d'articles scientifiques.

### **Aptitudes :**

- Savoir se situer dans des projets transversaux,
- Rigueur,
- Organisation,
- Recul,
- Travail en équipe

### **Formation**

Diplôme de niveau M2 en statistique ou en informatique

### **Expérience professionnelle**

Une expérience dans le domaine de la gestion de données est préférable

### **Contrat proposé**

Contrat à durée déterminée : temps plein de 12 mois renouvelable en tant qu'ingénieur d'étude

Rémunération : de 2138€ à 2800€ brut mensuel selon l'expérience par référence aux grilles de l'Inserm

Date de prise de fonction : 01/02/2021

**Pour postuler, merci d'envoyer CV et lettre de motivation à :**

CépiDc INSERM

[recrutement.cepidc@inserm.fr](mailto:recrutement.cepidc@inserm.fr)

Tel : 01 49 59 19 29