

# Sujet de stage de M2 statistique

## > LIEU DU STAGE

INSERM (Institut National de la Santé et de la Recherche Médicale)  
CépiDc (Centre d'épidémiologie sur les causes médicales de décès)  
80, rue du Général Leclerc, 94270 Le Kremlin-Bicêtre

## > INTITULE DU STAGE

Analyse des déterminants de la mortalité par cause dans la cohorte européenne EPIC

## > SUJET DU STAGE

### Domaine

Biostatistique, Epidémiologie

### Contexte

La cohorte européenne EPIC est l'une des plus volumineuses au monde, avec plus de 500 000 participants émanant de 10 pays. Elle a pour objectif l'analyse des facteurs liés au mode de vie, à la nutrition ou à des facteurs environnementaux sur la survenue de maladies chroniques comme le cancer.

Peu de travaux ont été jusqu'à présent menés à partir de cette cohorte sur l'analyse des déterminants de la mortalité par cause. Pourtant, notamment en raison de la simplicité de son interprétation et des efforts internationaux importants sur la standardisation des données de mortalité, ces indicateurs présentent un intérêt crucial en santé publique.

Cependant, plusieurs enjeux méthodologiques demeurent. En particulier, la prise en compte des causes multiples, permettant de ne pas considérer qu'une personne décède nécessairement d'une unique cause, a fait l'objet de développements récents au sein de l'équipe, et son application pratique sur des données internationales nécessite des stratégies statistiques à déterminer. Egalement, l'analyse de la mortalité par cause spécifique sur des données longitudinales nécessite la prise en compte du phénomène de concurrence des causes. Cette prise en compte est particulièrement nécessaire dans le cas d'une cohorte internationale où la distribution des causes de décès est très différente d'un pays à l'autre.

### Objectifs

L'objectif de ce stage consistera :

- en premier lieu à réaliser une analyse descriptive des données de mortalité de la cohorte EPIC, en cause initiale et par cause multiple, en comparant les distributions de mortalité par cause à celle des pays participants,
- en second lieu à appliquer des modèles de survie par cause spécifique, de type Age-Période-Cohorte, et à mesurer des gradients socio-économiques de mortalité,
- et de façon transversale, à appliquer des stratégies de pondération des causes multiples et de prise en compte des causes concurrentes, pour évaluer la sensibilité des résultats obtenus.

En cas de déroulement favorable, le stage pourra donner lieu à une suite (thèse de science) pour traiter plus en profondeur ces sujets.

### Matériel et méthodes

Les traitements porteront sur une extraction des bases de données d'EPIC, ainsi que sur les données de population et de mortalité en population générale.

Dans un premier temps et pour la prise en compte des causes multiples, des modèles de Cox adaptés seront appliqués, suivant la méthodologie développée dans :

- Moreno-Betancur M, Sadaoui H, Piffaretti C, Rey G. Survival analysis with multiple causes of death: Extending the competing risks model. *Epidemiology* 2017 (1):12-19.

Puis, dans une perspective exploratoire et pour la prise en compte des risques concurrents, des méthodes faisant intervenir des modèles avec IPCW (Inverse Probability of Censoring Weighting), pour estimer par exemple le SACE (Survivor Average Causal Effect), seront mises en œuvre.

Les traitements statistiques seront réalisés à l'aide des logiciels R et SAS

**Connaissances et aptitudes recherchées chez le stagiaire :**

*Connaissance des outils statistiques suivants :*

- *Analyse descriptive*
- *Modèles de régression*
- *Analyse de survie*

*Aptitude :*

- *Logiciels statistiques : R, SAS*
- *Anglais lu et écrit courant*

**> ENVIRONNEMENT DE LA MISSION****Intitulé, activité, compétences statistiques de l'unité d'accueil et du maître de stage :**

Le CépiDc (Centre d'épidémiologie sur les causes médicales de décès) est un service de l'Inserm (Institut national de la santé et de la recherche médicale), seul organisme public de recherche français entièrement dédié à la santé humaine.

En son sein, le CépiDc produit et diffuse la base nationale des causes médicales de décès, et développe des méthodologies statistiques d'exploitation des causes de décès.

Le stage sera encadré par Grégoire Rey, directeur du service, statisticien (diplômé de l'ISUP) et docteur en épidémiologie. Un comité de pilotage du stage se réunira tous les mois, il impliquera :

- Margarita Moreno-Betancur, docteur en biostatistique, spécialisée dans les méthodes d'analyse de survie par cause spécifique, Victorian Centre for Biostatistics (VICBiostat), Melbourne,
- Vivian Viallon, docteur en biostatistique, en charge de développements méthodologiques portant notamment sur l'exploitation des données de la cohorte EPIC, Centre International de Recherche sur le Cancer (CIRC), Lyon,
- Pietro Ferrari, Responsable du groupe de méthodologie nutritionnelle et de biostatistiques, Centre International de Recherche sur le Cancer (CIRC), Lyon.

**Ressources mises à la disposition du stagiaire :**

Le stagiaire disposera d'un bureau, d'un ordinateur puissant, et des logiciels SAS et R.  
La gratification du stage est d'environ 500€ / mois

**Début du stage :** février 2018

**Durée du stage :** 6 mois

**> PERSONNE(S) A CONTACTER**

**REY Grégoire, Directeur,**  
INSERM-CépiDc  
[gregoire.rey@inserm.fr](mailto:gregoire.rey@inserm.fr)  
**01 49 59 18 63**