

Stage de Master 2

Année universitaire 2019-2020

Santé publique France est l'agence nationale de santé publique française. Elle intervient au service de la santé des populations. Agence scientifique et d'expertise du champ sanitaire, elle a pour missions : (1) l'observation épidémiologique et la surveillance de l'état de santé des populations ; (2) la veille sur les risques sanitaires menaçant les populations ; (3) la promotion de la santé et la réduction des risques pour la santé ; (4) le développement de la prévention et de l'éducation pour la santé ; (5) la préparation et la réponse aux menaces, alertes et crises sanitaires ; (6) le lancement de l'alerte sanitaire.

■ Informations administratives

Titre	Modélisation multi-réponses de la dynamique des deux marqueurs TM et V3 chez les patients infectés par le VIH en France
Lieu de stage	Santé publique France. Direction Appui, Traitements et Analyses de données (DATA). 12 rue du Val d'Osne, Saint-Maurice.
Maitre de stage	Mme Cécile Sommen, biostatisticienne au sein de la direction Appui, Traitements et Analyse de données (DATA) et Julie Bertrand CR2 INSERM, IAME INSERM Université de Paris
Période de stage	Février-Juillet 2020 (6 mois)
Prérequis	Statistiques. Bonne maîtrise de R et de Monolix. La maîtrise du logiciel STAN serait un plus.

■ Sujet de stage

Contexte

Trente-cinq ans après la découverte du virus de l'immunodéficience humaine (VIH), l'épidémie se poursuit en France. Afin d'orienter les stratégies de prévention du VIH et de suivre leur impact, il est indispensable de connaître la dynamique de l'épidémie. L'indicateur permettant de rendre compte de l'évolution temporelle des infections est l'incidence. En France, depuis 2003, Santé publique France a mis en place un système de surveillance des nouveaux diagnostics VIH couplé à une surveillance virologique, permettant notamment de disposer des valeurs de deux marqueurs d'une infection récente (TM et V3) au moment du diagnostic de séropositivité.

Deux types de méthodes sont utilisés pour estimer l'incidence du VIH en France. Les méthodes basées sur le rétro-calcul utilisent les données de surveillance des nouveaux diagnostics VIH. Les méthodes basées sur les infections récentes utilisent la surveillance virologique des marqueurs V3 et TM. Ces dernières méthodes, utilisant des données plus informatives sur des infections récentes permettent une estimation de l'incidence plus précise dans les années les plus récentes. Cependant, elles nécessitent une connaissance de la dynamique des marqueurs depuis l'infection.

Données PRIMO ANRS-C06

Pour modéliser la dynamique des marqueurs TM et V3, les données de la cohorte PRIMO-ANRS C06 seront utilisées. Cette cohorte inclut plus de 250 individus pendant la phase de primo-infection à

partir de novembre 1996 dans 66 hôpitaux français. Les sujets de la base sont naïfs de traitements antirétroviraux à leur inclusion dans la cohorte. Les patients ont été examinés cliniquement et biologiquement au mois M1, M2, M3 et M6 et par la suite, tous les 6 mois avec des mesures de concentration des marqueurs TM et V3 à chaque visite. Plus précisément, les marqueurs ont été recueillis en utilisant des échantillons de sang sec sur papier buvard. Ces marqueurs ont été choisis de manière à ce qu'ils soient reconnus peu de temps après la séroconversion pour rendre possible la distinction entre infection récente et infection non récente.

Méthode d'estimation des paramètres du modèle

En 2009, une première modélisation conjointe de l'évolution des marqueurs TM et V3 a été effectuée en utilisant un modèle non linéaire mixte multi-réponses avec mouvement Brownien. En 2010, une modélisation du marqueur TM uniquement a été effectuée en utilisant un modèle non linéaire mixte et le package SAEMIX du logiciel R pour l'estimation des paramètres.

Objectifs de l'étude

L'objectif principal de ce stage est la modélisation multi-réponses des marqueurs TM et V3, en utilisant deux approches l'une fréquentiste et l'autre bayésienne, en utilisant :

1. L'algorithme SAEM dans le logiciel Monolix pour l'approche fréquentiste
2. L'algorithme Hamiltonien de Monte Carlo dans le logiciel Stan pour l'approche bayésienne

Une comparaison des deux approches en termes d'estimations obtenues mais aussi en termes de difficultés d'implémentation, de temps de calcul sera effectuée et discutée.

Analyses statistiques

Les méthodes statistiques envisagées seront une méthode fréquentiste et une méthode bayésienne permettant l'estimation de modèles mixtes non linéaires multi-réponses pour deux marqueurs quantitatifs. Le candidat doit être capable de comprendre les programmes d'estimation existants, de les modifier et de programmer de nouveaux si nécessaires, notamment avec le logiciel Monolix et le logiciel Stan. Une bonne connaissance du logiciel R est également nécessaire.

Les étapes du stage à réaliser

- Revue de la littérature, compréhension des méthodes existantes pour l'estimation de la dynamique des marqueurs TM et V3 et compréhension à mettre en place pendant le stage
- Implémentation des deux méthodes dans les différents logiciels
- Comparaison des deux approches et discussion des résultats