Evaluation

#### Juliette Murris

#### Janvier 2023

En binôme, vous répondrez aux différentes questions posées et rapporterez vos travaux sous la forme d’une présentation.

La présentation devra durer 10 minutes, suivie de 10 minutes de questions. Elle aura lieu le lundi 30 janvier 2022.

Les sujets

L’objectif est d’évaluer l’association traitement/événement pour une population d’intérêt. L’événement ne peut pas être le décès. N’oubliez pas de justifier vos choix et de vérifier vos hypothèses. Le décès peut être l'événement compétitif mais pas principal

En binôme, choisissez l’un des sujets suivants:

Sujet 1 - C’est vous qui choisissez!

Sélectionnez un traitement/un événement/une population/une maladie qui vous questionne

Identifiez une source de données qui permet de répondre à cet enjeu

Avec au minimum 2 événements compétitifs (l’évènement 1 ne peut pas se produire avec l’évènement 2), 3 facteurs et 1 facteur hiérarchique (permet de clusterisé les patients ensemble)

Où trouver des données?

Données disponibles dans certaines publications

Plateformes open data : data.europa, re3data, etc.

Sujet 2 - Approche par la recherche: simulez des données, à vos claviers!

A l’aide d’un package de simulation sous R, simuler des données de survie avec au minimum 2 événements compétitifs, 3 facteurs et 1 facteur hiérarchique

Positionnez-vous en vraie vie: donnez du sens à vos simulations en les nommant. Par exemple: il s’agit de simuler des données sur l’effet de la chimiothérapie sur la rechute auprès des patients atteintes de cancer du sein HER2+, en prenant comme facteur le site de métastases, l’âge, la classification TNM, etc.

Répondez ensuite aux questions suivantes:

1. Décrire la base de données en fonction du traitement (test stat et visualisation)
2. Tracer les courbes de survie de chaque événement
3. Quantifier l’effet du traitement sur chaque événement, ajusté sur les facteurs
4. Prendre en compte l’effet hiérarchique
5. Quelles sont les limites de ces analyses?
6. Faire les mêmes analyses en risques compétitifs: Tracer incidence cumulée et utiliser le modèle de Fine and Gray
7. Commenter
8. Choisir une méthode de validation des données : traint-test split, k-folds, etc.
9. Choisir une méthode de ML afin de classer les patients selon l’événement d’intérêt
10. Evaluer les performances du modèle
11. Identifier les variables qui semblent le mieux permettre cette classification
12. Répondre à la question du projet
13. Enoncer les difficultés rencontrées et les solutions apportées

Contraintes data set :

1. Médical
2. Au moins deux évent qui sont en compétition (mort et survenue de la maladie/ soit 1 soit2, deux event pas forcement depend et pas avoir les deux en même temps) [survie ne dois pas être définie par la mort]
3. Facteur Hierarchique (on peut le crée) rassembler les patients dans des groupes distincts (age, sexe, pas une variable déjà définie pas d’effet centre, region, pays)
4. 3 variables