Offre de stage de Master 2 en Biostatistiques

Intitulé : Émulation d'essai cible dans les néoplasies myéloprolifératives : comparaison d'estimateurs causaux

Formation : Master 2 en Biostatistiques, Statistiques ou Épidémiologie

Lieu : CHU Angers, Département des Sciences de la Donnée, Unité Biostatistiques et Méthodologie

Durée : 6 mois (début souhaité entre janvier et mars 2027) en présentiel

Contexte

Le traitement des patients atteints de néoplasie myéloproliférative (NMP) repose sur les agents antithrombotiques et cytoréducteurs, afin de prévenir le risque de thrombose et la transformation en myélofibrose ou en leucémie aiguë. Il n'est cependant pas toujours facile de déterminer l'intensité optimale du traitement pour chaque patient et de limiter la toxicité médicamenteuse (hémorragies, cytopénies, etc.).

Les essais randomisés, qui apportent le niveau de preuve le plus élevé pour la pratique clinique, sont rares dans ces pathologies. Les événements d'intérêt (thrombose, hémorragie, transformation) sont peu fréquents et peuvent survenir 5 à 10 ans après la randomisation : il faudrait de larges effectifs et des essais très longs.

L'émulation d'essai cible (*target trial emulation*) offre une alternative : on spécifie d'abord le protocole de l'essai qu'on aurait voulu mener (critères d'éligibilité, stratégies de traitement, temps zéro, critère de jugement, contraste causal), puis on l'émule avec des données observationnelles, ce qui limite des biais « auto-infligés » comme le biais de temps immortel [1]. Les hypothèses causales sont formalisées par des graphes acycliques dirigés (DAG). La cohorte FIMBANK, qui regroupe 5057 patients atteints de NMP, environ 19 000 consultations de suivi et une durée médiane de suivi de 6 ans, constitue une base adaptée pour émuler des essais afin d'orienter la pratique clinique là où un essai randomisé est difficile à mener.

Lorsque les facteurs de confusion évoluent au cours du temps et sont eux-mêmes affectés par les traitements antérieurs, il est nécessaire de recourir à des méthodes adaptées à l'estimation causale longitudinale (g-methods), notamment les Marginal Structural Models (MSM) estimés par Inverse Probability Weighting (IPW), la parametric g-formula et les estimateurs doublement robustes tels que l'Augmented Inverse Probability Weighting (AIPW) et le Targeted Maximum Likelihood Estimation (TMLE). Ces approches peuvent également intégrer des méthodes de machine learning [2]. Leur comportement lorsque les événements sont rares et les visites peu fréquentes reste encore à documenter, d'où une étude de simulation structurée selon le cadre ADEMP, calibrée sur FIMBANK.

Objectif

Estimer l'effet causal de stratégies thérapeutiques dans les NMP par émulation d'essai cible à partir de la cohorte FIMBANK. Le stage combine deux volets : une étude de simulation pour identifier les estimateurs causaux les plus fiables (MSM estimés par IPW, parametric g-formula, AIPW/TMLE) lorsque les événements sont rares et les facteurs de confusion varient dans le temps, puis l'application de ces méthodes à FIMBANK pour estimer l'effet causal d'une stratégie thérapeutique.

Missions

1. **Cadrage :** revue de la littérature, rédaction du protocole de l'essai cible (check-list TARGET) et du DAG des hypothèses causales, validés avec les cliniciens.

2. **Simulation :** construction d'un simulateur avec vérité causale connue, fondé sur le DAG et calibré sur FIMBANK, et rédaction du protocole de simulation (cadre ADEMP).

3. **Implémentation des estimateurs causaux :** mise en œuvre de Marginal Structural Models (MSM) estimés par Inverse Probability Weighting (IPW), de la parametric g-formula et des estimateurs doublement robustes AIPW et TMLE, avec des modèles paramétriques puis des méthodes de machine learning, notamment le Super Learner.

4. **Étude comparative :** scénarios de difficulté croissante (événements rares, mauvaise spécification des modèles, positivité faible c'est-à-dire peu de patients comparables entre groupes, censure informative) et évaluation du biais, de la variance, de l'EQM et de la couverture.

5. **Application à FIMBANK :** émulation d'un essai dans la cohorte, avec analyses de sensibilité.

Compétences souhaitées

- Bases solides en statistiques, modèles de régression et analyse de survie. Des compétences en inférence causale et en apprentissage automatique seront appréciées.
- Bonne maîtrise de la programmation en R, et goût pour la simulation.
- Bonnes compétences en anglais (lu et écrit).
- Connaissance de la recherche clinique appréciée.

Encadrement : LU Yunjiao (PhD), MOREAU Clémence (PhD) et RIOU Jérémie (PU-PH).

Gratification : gratification légale de stage, selon les modalités en vigueur au CHU d'Angers.

Modalités de candidature : CV, lettre de motivation et relevés de notes si disponibles, à envoyer à Clemence.Moreau@chu-angers.fr.

Références

1. Hernán MA, Robins JM. Using big data to emulate a target trial when a randomized trial is not available. *Am J Epidemiol*. 2016;183(8):758-764.

2. Hernán MA, Robins JM. *Causal Inference: What If*. Chapman & Hall/CRC; 2020.