

Apports de la modélisation causale dans l'évaluation des immunothérapies à partir de données observationnelles

Les lymphocytes T du système immunitaire ont la capacité d'éliminer les cellules cancéreuses. Or, il existe des protéines qui freinent leur agressivité : les cellules cancéreuses exprimant ces protéines peuvent échapper à leur destruction par les lymphocytes T. De nouvelles stratégies d'immunothérapie ont été proposées ; elles sont basées sur des molécules qui inhibent certains de ces freins, les protéines CTLA-4 ou PD-1, pour rendre le système immunitaire plus efficace contre les cellules cancéreuses. Ces immunothérapies ont été évaluées dans des essais thérapeutiques de phases I à III et, notamment, les anti-PD1/anti-PDL1 ont prouvé leur efficacité dans différentes pathologies, le mélanome métastatique, les cancers du rein métastatiques, les cancers du poumon.

Tous les patients ne répondent pas à ces nouvelles immunothérapies avec la même intensité. Afin de pouvoir identifier les patients bénéficiant de ces immunothérapies, on mesure la présence de nombreux marqueurs immunologiques exprimés par les tumeurs à l'instauration du traitement et leur évolution sous traitement. Les anti-CTLA4 et les anti-PD1 ciblant des mécanismes immunologiques distincts, leur action est complémentaire. Leur administration simultanée conduit à un niveau de toxicité élevé ; on ne sait pas encore s'il est préférable de traiter d'abord par anti-CTLA4 puis par anti-PD1/anti-PDL1 ou l'inverse. Le schéma optimal de traitement par immunothérapie n'a pas encore été déterminé : fréquence des injections, traitement en continu versus phase d'induction avec ou sans maintenance versus traitement on/off.

En situation observationnelle, l'évaluation de l'effet d'un traitement sur la survenue d'un événement implique de contrôler les facteurs de confusion. En l'absence de tirage au sort du traitement, traités et non traités peuvent ne pas être comparables sur leur risque de base de l'événement (absence d'interchangeabilité). L'interchangeabilité entre les traités et les non traités n'est pas garantie : leur risque de base de décès peut différer. Classiquement, les modèles multivariés ajustés sur les facteurs de confusion visent à identifier les facteurs indépendants de survenue d'un événement. Toutefois, dans certaines situations, quand le facteur est une covariable dépendante du temps ou quand les relations ordonnées entre les facteurs sont inconnues, les méthodes classiques échouent à contrôler les biais de sélection.

Les méthodes d'inférence causale sur données observationnelles [1- 3] ont été introduites dans ce contexte. Après avoir identifié les sources de biais de sélection par la construction de diagrammes acycliques dirigés, des modélisations causales visent à atteindre l'interchangeabilité conditionnelle entre exposés et non exposés. Parmi elles, la pondération par l'inverse du score de propension d'exposition au facteur étudié construit à partir des données observationnelles une pseudo-population où exposés et non exposés sont interchangeables (conditionnellement). Les conditions d'application de cette méthode sont l'interchangeabilité conditionnelle qui requiert l'exhaustivité du recueil des facteurs de confusion, la positivité (existence d'exposés et de non exposés pour la même combinaison de facteurs de confusion) et la cohérence (en anglais, consistency). D'autres modélisations causales ont été proposées dont la « G-Formula paramétrique » et la « G-estimation » [4,5] notamment pour être employées quand les conditions d'application de la pondération par l'inverse du score de propension ne sont pas remplies. Dans

le contexte d'exposition à des facteurs multidimensionnels (ex : marqueurs immunologiques), des méthodes causales dédiées aux données de grande dimension [6,7] ont été proposées.

La thèse a pour objectifs de développer des approches causales et de les appliquer aux questions posées par l'évaluation des immunothérapies.

- Axe 1 : Inférence causale dans les analyses de marqueurs immunologiques
 - Méthodes: Théorie des graphes (Directed Acyclic Graphs (DAGs)), adaptation du Causal Stability Ranking (CStaR) dans le contexte où les biomarqueurs sont évalués à l'initiation du traitement et sous traitement (effet des valeurs de base et des pentes d'évolution sous traitement)
- Axe 2 : Identifier la meilleure séquence thérapeutique entre anti-CTLA4 et anti-PD1 ou anti-PD1 puis anti-CTLA4 à partir de données observationnelles
 - Méthodes: DAGs, Adaptation de la Pondération inverse par le score de propension (IPW), de la G-formula, de la G-estimation à l'évaluation de deux séquences de deux traitements, le second étant initié après l'échec du premier
- Axe 3 : Comparer l'efficacité et la toxicité de différents schémas thérapeutiques d'immunothérapie à partir de données observationnelles
 - Méthodes: DAG, Adaptation des méthodes IPW, G-formula, G-estimation à la comparaison de schémas thérapeutiques faisant varier la fréquence des injections

Bibliographie

1. Hernán MA, Lanoy E, Costagliola D, Robins JM. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin Pharmacol Toxicol*. 2006 Mar;98(3):237-42.
2. Hernán MA, McAdams M, McGrath N, Lanoy E, Costagliola D. Observation plans in longitudinal studies with time-varying treatments. *Stat Methods Med Res*. 2009
3. Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernán MA. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *Int J Biostat*. 2010;6(2)
4. Westreich D, Cole SR, Young JG, Palella F, Tien PC, Kingsley L, Gange SJ, Hernán MA. The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident AIDS or death. *Stat Med*. 2012
5. Picciotto S, Hernán MA, Page JH, Young JG, Robins JM. Structural Nested Cumulative Failure Time Models to Estimate the Effects of Interventions. *J AmStat Assoc*. 2012;107(499).
6. Stekhoven DJ, Moraes I, Sveinbjörnsson G, Hennig L, Maathuis MH, Bühlmann P. Causal stability ranking. *Bioinformatics*. 2012 28(21):2819-23.
7. Ternès N, Arnedos M, Koscielny S, Michiels S, Lanoy E. Statistical methods applied to omics data: predicting response to neoadjuvant therapy in breast cancer. *Curr Opin Oncol*. 2014 Nov;26(6):576-83.

Directeurs de thèse

- Lanoy Emilie (PhD)
- Michiels Stefan (HDR)

Équipe 2 - Méthodologie et épidémiologie clinique en oncologie moléculaire (Responsable : Stefan Michiels) du Centre de recherche en épidémiologie et santé des populations (CESP)-Inserm, unité mixte de recherche 1018

Service de Biostatistique et Epidémiologie (dirigé par le Dr Ellen Benhamou)

Gustave-Roussy, 114, rue Edouard Vaillant, 94805 Villejuif

<http://cesp.inserm.fr/equipe-2.html>

Profil souhaité du candidat

- Titulaire d'un master 2 de Santé Publique (Biostatistique, épidémiologie ou recherche clinique) ou équivalent
- Solides compétences en méthodologie statistique fréquentiste, analyse de survie multivariée
- Compréhension écrite et orale en anglais
- Programmation en R et/ou en SAS
- Lecture critique d'articles
- Capacités de travail en équipe et en autonomie
- Intérêt pour les questions cliniques, en oncologie notamment
- Connaissance de la méthodologie des essais thérapeutiques en oncologie serait un plus

Les candidatures sont à adresser à emilie.lanoy@gustaveroussy.fr.