

# Méthodes de sélection de variables dans le cadre des modèles à risques compétitifs : Application à la cancérologie

**Responsable :** Thomas Filleron

**Tél :** +33 (0) 531 155 865

**Mél :** filleron.thomas@iuct-oncopole.fr

**Lieu :** IUCT-O. Institut Claudius Regaud. 1 avenue Irène Joliot-Curie. 31059 TOULOUSE Cedex 9

## Contexte Scientifique :

Afin de guider le clinicien dans sa démarche diagnostique et/ou thérapeutique, des modèles mathématiques combinant plusieurs variables clinico-pathologiques sont proposés dans la littérature avec pour objectif de prédire la survenue d'un événement d'intérêt. Les nomogrammes permettent d'obtenir des prévisions de façon individuelle pour chaque patient. En 2010, un nomogramme a été développé pour prédire la probabilité de survenue des métastases cérébrales chez les patientes traitées pour un cancer du sein métastatique (Graesslin, 2010). Ce nomogramme a été développé sur une base de données du MD Anderson cancer- et - a fait l'objet d'une validation externe dans une population de patients du Cross Cancer Institute (Edmonton, Alberta, Canada). Récemment, ce nomogramme a été appliqué à une population de patients de l'Institut Claudius Regaud (Genre, 2017). Nous avons montré que la discrimination était similaire à celle obtenue sur la première validation et que la calibration était bonne. Cependant deux principales critiques sur ce nomogramme sont mises en évidence:

- il a été développé en considérant un critère de jugement binaire qui ne prend pas en compte le temps jusqu'à l'apparition des métastases cérébrales
- il ne prend pas en compte les risques compétitifs. En effet, les patientes sont à risque de métastase cérébrale (événement d'intérêt) mais elles sont également à risque de décéder avant l'apparition de métastase cérébrale (événement compétitif).

L'objectif principal de ce stage va être d'étudier les différentes méthodes de sélection de variables dans le cadre des modèles à risques compétitifs. Ces différentes méthodologies seront appliquées à ce jeu de données afin de proposer un score pronostique visant à identifier les patientes traitées pour un cancer du sein métastatique qui présenteront des métastases cérébrales au cours du temps.

Le modèle semi-paramétrique de Fine & Gray a une formulation similaire au modèle de Cox mais pour la fonction de risque associée à la fonction d'incidence cumulée («subdistribution Hazard»). Pour cette méthode, qui est devenue la référence pour l'analyse multivariée des risques compétitifs, différentes méthodes de sélection de variables ont été proposées dans la littérature. La première stratégie consiste à incorporer les covariables à l'aide d'une procédure pas à pas descendante ou ascendante et différents critères de sélection peuvent être utilisés (Kuk, Stat Med 2013). D'autres approches basées sur les méthodes de pénalisation, de boosting ou sur les forêts aléatoires ont également été proposées dans la littérature (Fu, 2016 ; Binder, 2009 ; Ishwaran, 2014). Après avoir pris connaissance de ces différentes approches, elles seront appliquées à une base de données de patientes traitées pour un cancer du sein métastatique. Pour chaque approche, un score continu sera proposé. Comme il est souvent utile de disposer de valeurs « seuil » pour prendre une décision clinique, il sera dichotomisé pour proposer deux groupes de risque. Les performances de chaque score et des groupes de risque associés seront évaluées à l'aide d'une validation interne.

**Encadrement / Information Pratique :** Le stagiaire bénéficiera d'un encadrement sur le plan méthodologique et sur le plan clinique.

## Profil recherché :

**Formation :** Master 2 en Biostatistique / Dernière année d'école d'ingénieurs

**Prérequis :** Bonnes connaissances en biostatistique (Modélisation), analyse de survie, risques compétitifs et bonne connaissance de la programmation sous R et/ou Stata. Un caractère dynamique, entreprenant et motivé, ainsi qu'un esprit multidisciplinaire seront des atouts.

## Bibliographie

- Fine JP, Gray RJ: A proportional hazards model for the subdistribution of a competing risk. Journal of the American Statistical Association 94:496-509, 1999
- Fu Z, Parikh CR, Zhou B. Penalized variable selection in competing risks regression. Lifetime Data Anal. 2016 Mar 26;
- Binder H, Allignol A, Schumacher M, Beyersmann J. Boosting for high-dimensional time-to-event data with competing risks. Bioinformatics. 2009 Apr 1;25(7):890-6.
- Ishwaran H, Gerds TA, Kogalur UB, Moore RD, Gange SJ, Lau BM. Random survival forests for competing risks. Biostatistics. 2014;15(4):757-773.
- Kuk D, Varadhan R. Model selection in competing risks regression. Stat Med. 2013 Aug 15;32(18):3077-8
- Genre L, Roché H, Varela L, Kanoun D, Ouali M, Filleron T, Dalenc F. External validation of a published nomogram for prediction of brain metastasis in patients with extra-cerebral metastatic breast cancer and risk regression analysis. Eur J Cancer. 2017 Feb;72:200-209.
- Graesslin O, Abdulkarim BS, Coutant C, Huguet F, Gabos Z, Hsu L, Marpeau O, Uzan S, Pusztai L, Strom EA, Hortobagyi GN, Rouzier R, Ibrahim NK.
- Nomogram to predict subsequent brain metastasis in patients with metastatic breast cancer. J Clin Oncol. 2010 Apr 20;28(12):2032-7.