INSERM - Unité 897
Centre de Recherche Epidémiologie et Biostatistique

Sujet de stage Master II en Biostatistique

Modélisation de données répétées ordinales en présence de sortie d'étude informative

Co-direction :
Cécile Proust-Lima et Loïc Ferrer
Cecile.proust-lima@inserm.fr

Il est très fréquent en essais cliniques ou en études observationnelles de devoir analyser des données ordinales collectées de façon répétée. Des exemples incluent des échelles de douleur, de dépression ou de qualité de vie. Ces indicateurs, que l'on peut souvent retrouver dans la littérature sous l'intitulé PRO pour « patient reported outcomes », sont souvent analysés de façon longitudinale avec des modèles linéaires mixtes sous l'hypothèse qu'ils sont continus Gaussiens dès lors qu'ils contiennent un certain nombre de niveaux différents. Mais, même avec un grand nombre de niveaux différents, il a été montré que faire cette hypothèse pouvait induire des biais importants sur les estimations des paramètres d'intérêt (effet de variables explicatives) (Proust-Lima et al., 2011). Des modèles adaptés pour des données ordinales répétées sont donc à privilégier. Ces modèles sont par exemple les modèles probit cumulés ou modèles logistiques à côtes proportionnelles, étendus au cas longitudinal par l'inclusion d'effets aléatoires (Tutz et Hennevogl, 1996 ; Hedeker et Gibbons, 1994).

Dans les essais cliniques ou les études observationnelles, le suivi des sujets est souvent interrompu par l'occurrence d'événements cliniques (diagnostic, rechute, décès). Ces événements induisent souvent une sortie d'étude informative qu'il est nécessaire de prendre en compte pour éviter des biais dans l'estimation des trajectoires de la variable d'intérêt. La prise en compte de sortie d'étude informative en analyse longitudinale peut se faire par la modélisation simultanée (dite conjointe) des données répétées de la variable d'intérêt et du processus de sortie d'étude. L'approche de modélisation conjointe s'est très développée ces dernières années en biostatistique (Rizopoulos, 2012 ; Ibrahim et al., 2010) mais l'application aux données répétées ordinales est encore peu courante du fait de la complexité inférentielle induite et du manque de logiciel statistique approprié (Li et al., 2010).

L'objectif de ce stage est de proposer et implémenter un modèle conjoint basé sur une nouvelle technique inférentielle qui permettra d'analyser des données répétées ordinales en présence d'une sortie d'étude informative. Ce modèle permettra de prendre en compte possiblement une ou plusieurs variables ordinales répétées corrélées entre elles (Liu et Hedeker, 2006).

L'implémentation du modèle se fera de façon préférentielle sous R avec la possibilité d'associer un autre langage de programmation (e.g., C++).

La méthodologie sera appliquée à des données de qualité de vie issues d'un essai clinique en cancérologie.

Compétences requises :
Très bonnes connaissances en biostatistique, notamment les méthodes d'analyse de données longitudinales : connaissance du modèle linéaire mixte primordiale.
Bonne connaissance des techniques algorithmiques et d'un langage de programmation.
Maitrise du logiciel R.

Structure d'accueil :
Centre de Recherche INSERM U897 – Epidémiologie-Biostatistique
ISPED Université de Bordeaux
146, rue Léo-Saignat
33076 Bordeaux Cedex
http://biostat.isped.u-bordeaux2.fr

Durée du stage : 4-6 mois
Période du stage : début du stage entre janvier et avril 2016
Gratification de stage : selon les grilles prévues par l'Inserm

Contact : adresser une lettre de motivation et un CV à cecile.proust-lima@inserm.fr

Références :

Hedeker, D., & Gibbons, R. D. (1994). A random-effects ordinal regression model for multilevel analysis. *Biometrics*, *50*(4), 933-944.

Ibrahim, J. G., Chu, H., & Chen, L. M. (2010). Basic Concepts and Methods for Joint Models of Longitudinal and Survival Data. *Journal of Clinical Oncology*, *28*(16), 2796-2801.

Li, N., Elashoff, R. M., Li, G., & Saver, J. (2010). Joint modeling of longitudinal ordinal data and competing risks survival times and analysis of the NINDS rt-PA stroke trial. *Statistics in Medicine*, *29*(5), 546–557.

Liu, L. C., & Hedeker, D. (2006). A mixed-effects regression model for longitudinal multivariate ordinal data. *Biometrics*, *62*(1), 261-8.

Proust-Lima, C., Dartigues, J. F., & Jacqmin-Gadda, H. (2011). Misuse of the linear mixed model when evaluating risk factors of cognitive decline. *Am J Epidemiol*, *174*(9), 1077-88.

Rizopoulos, D. (2012). Joint Models for Longitudinal and Time-to-Event Data: With Applications in R. *Chapman & Hall/CRC Biostatistics Series*.

Tutz, G., & Hennevogl, W. (1996). Random Effects in Ordinal Regression Models. *Comput. Stat. Data Anal.*, *22*(5), 537–557.