

# MODÉLISATION DE L'EFFET DE VARIABLES EXPLICATIVES SUR DES INDICATEURS DE SANTÉ : APPROCHE PAR PSEUDO-VALEURS

Camille Sabathé <sup>1,\*</sup> & Pierre Joly <sup>1</sup>

<sup>1</sup> *Univ. Bordeaux, INSERM, Bordeaux Population Health, U1219, Bordeaux - France*

\* *camille.sabathe@u-bordeaux.fr*

L'objectif de ce travail est d'étudier l'effet de variables explicatives sur des indicateurs de santé tels que la probabilité de devenir dément ou la probabilité de rester non dément et vivant. Ces indicateurs de santé dépendent à la fois du risque de maladie et à la fois du risque de décès. Les méthodes utilisées usuellement modélisent le risque instantané de tomber malade (modèle à risque proportionnel cause-spécifique) ou sont difficilement interprétable (modèle de Fine et Gray (1999)).

Une approche par pseudo-valeurs issues d'estimateurs non paramétriques et développée par Andersen *et al.* (2003) permet de modéliser directement l'effet de variables explicatives sur ces probabilités dans le cadre de données censurées à droite. Soit  $\theta$  une quantité d'intérêt et  $Y_i$  la pseudo-valeur du sujet  $i$ . La pseudo-valeur est définie par :

$$Y_i = n \times \hat{\theta} - (n - 1) \times \hat{\theta}^{-i} \quad (1)$$

avec  $\hat{\theta}$  l'estimation d'une quantité à partir de l'échantillon de  $n$  sujets,  $\hat{\theta}^{-i}$  la même estimation à partir de l'échantillon sans le sujet  $i$  et  $n$  le nombre de sujets de l'échantillon. Après avoir été calculées, les pseudo-valeurs sont utilisées comme variable d'intérêt dans un modèle GLM pour regarder l'impact des variables explicatives sur les indicateurs de santé.

Ce travail consiste précisément à étendre l'approche par pseudo-valeurs à des données censurées par intervalle pour pouvoir être appliquée aux données de cohorte. La maladie est observée en temps discret (à chaque visite de suivi de la cohorte) alors qu'elle apparaît en temps continu. Cette censure par intervalle entraîne une inconnue sur la date précise de début de la pathologie et une inconnue sur les trajectoires des individus. En effet, les individus qui décèdent alors qu'ils étaient non déments à leur dernière visite ne peuvent être observés déments.

Pour utiliser l'approche par pseudo-valeurs sur des données censurées par intervalle, les indicateurs de santé sont estimés à partir d'un estimateur du maximum de vraisemblance pénalisée approché par des splines (Joly *et al.*, 2002).

Les performances de cette approche ont été testées par des simulations. Une application de cette méthode est faite sur les données de la cohorte Paquid (Dartigues *et al.*, 1992).

## Références

- P. K. ANDERSEN, J. P. KLEIN et S. ROSTHOJ : Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, 90(1):15–27, 2003.
- J. F. DARTIGUES, M. GAGNON, P. BARBERGER-GATEAU, L. LETENNEUR, D. COMMENGES, C. SAUVEL, P. MICHEL et R. SALAMON : The Paquid epidemiological program on brain ageing. *Neuroepidemiology*, 11 Suppl 1:14–18, 1992.
- J. P. FINE et R. J. GRAY : A Proportional Hazards Model for the Subdistribution of a Competing Risk. *Journal of the American Statistical Association*, 1999.
- P. JOLY, D. COMMENGES, C. HELMER et L. LETENNEUR : A penalized likelihood approach for an illness-death model with interval-censored data : application to age-specific incidence of dementia. *Biostatistics*, 3(3):433–443, 2002.