

MODÈLE ILLNESS-DEATH STRATIFIÉ POUR DONNÉES GROUPÉES

Pierre Joly^{1,2}, Célia Touraine^{3,4}, Catherine Helmer^{1,2} & Hélène Jacqmin-Gadda ^{1,2}

¹ *Université de Bordeaux, Isped, Bordeaux, France*

² *Inserm U897, Bordeaux, France*

³ *INSERM, UMR 912 (SESSTIM), 13006, Marseille, France*

⁴ *Université d'Aix-Marseille, UMR_S912, IRD, 13006, Marseille, France*

Résumé. L'objectif de ce travail est d'étudier les facteurs de risque de démence à partir de données de cohorte et de calculer des espérances de vie. La population d'étude est constituée de sujets âgés de plus de 65 ans et vivants à domicile dans 75 communes de Gironde et de Dordogne. Les sujets, initialement non déments, ont été suivis plus de 20 ans. L'évènement étudié est l'apparition d'une démence et le décès est un risque concurrent. De plus l'âge d'apparition de la démence est censuré par intervalles car les sujets sont diagnostiqués durant des visites de suivis planifiées. Enfin, il est admis que les taux de décès en fonction de l'âge diffèrent entre les hommes et les femmes.

Pour prendre en compte ces contraintes nous utilisons un modèle Illness-death irréversible pour données tronquées à gauche et censurées par intervalles que nous généralisons en le stratifiant sur le sexe. Les facteurs de risque sont étudiés grâce à un modèle de régression à intensités de transition proportionnelles et nous avons rajouté au modèle des effets aléatoires pour tenir compte du regroupement par communes des sujets.

Mots-clés. Modèle Illness-death ; données groupées ; modèle à effets aléatoires ; censure par intervalles ; espérance de vie

Abstract. The objective of this work is to study risk factors of dementia from cohort data and to estimated life expectancies. The population consists of subjects aged 65 years and older living in 75 municipalities of Southwestern France. The subjects where initially non-demented and have been followed for dementia over more than 20 years. The event of interest is dementia and death without dementia is a competing risk. We had to deal with interval censoring because subjects were regularly screened and diagnosed for dementia at planned visits after entry. For subjects diagnosed with dementia, we only know that the exact age at disease onset was between the diagnostic visit and the previous one. Concerning deceased subjects without being diagnosed as demented at their last visit, we cannot ignore their risk of having been diseased between their last visit and death. Age was chosen as the basic time-scale so we had to deal with left-truncation. Age-specific incidence of dementia and mortality rates differs according to gender.

To deal with all these issues we propose to use an illness-death model stratified on gender and including random effects to take into account the cluster structure. Proportional transition intensities models are used to assess the impact of individual factors on

each transition. The illness-death model allows individuals to move from the “healthy” state to the “dead” state either directly or having been in the “diseased” state before. For inference, right and interval censoring as well as left truncation have to be dealt with.

Keywords. Illness-death model; grouped data; Random effects; Interval censoring; life expectancy.

1 Motivations épidémiologiques

L’objectif est d’étudier les facteurs de risque de développer une démence ainsi que de calculer des espérances de vie, pour des individus déments ou non déments, à partir de données issues d’une cohorte. Dans cette cohorte les sujets ont été inclus s’ils étaient âgés de 65 ans ou plus et s’ils étaient non déments et vivaient à domicile. Les sujets ont été tirés au sort dans 75 communes de Gironde et de Dordogne. Le nombre de participants par commune varie entre 15 et 724. Ensuite les sujets ont été revus à des visites environ 1, 3, 5, 8, 10, 13, 15, 17 et 20 ans après l’entrée dans la cohorte. Du fait de l’âge élevé des sujets le risque de développer une démence est en concurrence avec le risque de décéder. Du fait du suivi des sujets si un sujet est diagnostiqué dément à une visite, son âge d’apparition de la démence est compris entre son âge à la dernière visite sans démence et son âge au diagnostic. De plus, du fait de la concurrence avec le décès, des sujets ont pu développer une démence entre leur dernière visite et leur décès sans qu’ils aient pu être diagnostiqués déments. Les données sont donc censurées par intervalles et aussi tronquées à gauche car les sujets étaient vivants et non déments à l’entrée dans l’étude. Enfin, il est admis que les taux de décès en fonction de l’âge diffèrent entre les hommes et les femmes. Nous sommes donc en présence d’un risque concurrent différent d’un sexe à l’autre avec des données censurées par intervalles, tronquées à gauche et groupées par communes.

2 Le modèle

Pour prendre en compte les spécificités des données pour l’étude des facteurs de risque de démence nous avons choisi d’utiliser un modèle “Illness-death” présenté sur la figure 1. Avec ce modèle, les sujets peuvent passer de l’état vivant non dément (l’état 0 dans la figure 1) à un état dément (l’état numéroté 1) et à un état décédé (l’état 2) soit directement de l’état 0, pour les sujets décédant avant de devenir déments, soit depuis l’état 1 pour les sujets qui décèdent en étant déments.

Les fonctions $\alpha_{01}(t)$, $\alpha_{02}(t)$ et $\alpha_{12}(t)$ sont les intensités de transition. Pour prendre en compte des facteurs de risque nous utilisons un modèle de régression à intensités de transition proportionnelles (Joly et al, 2002). Pour un sujet j et pour la transition $k \rightarrow l$,

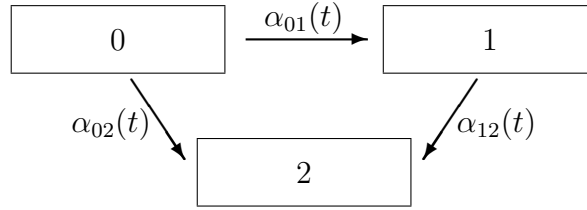


Figure 1: Le modèle Illness-death

$(k, l) \in \{(0, 1), (0, 2), (1, 2)\}$ on a :

$$\alpha_{kl}(t|\mathbf{z}_{kl,j}, \boldsymbol{\beta}_{kl}) = \alpha_{kl}^0(t) e^{\boldsymbol{\beta}_{kl}^T \mathbf{z}_{kl,j}}$$

où $\alpha_{kl}^0(\cdot)$ est l'intensité de transition de base et $\mathbf{z}_{kl,j}$ le vecteur des variables explicatives.

Pour introduire dans le modèle un éventuel effet groupe (commune dans notre application) sur les intensités de transition nous utilisons des effets aléatoires (voir Duchateau et Janssen (2008) et Joly et al (2012)). Les intensités de transition de base sont différentes entre les hommes et les femmes. Dans ce cas, pour un sujet j du groupe i avec un sexe s_{ij} , les intensités de transition sont :

$$\alpha_{01}(t|\mathbf{z}_{01,ij}, \boldsymbol{\beta}_{01}, u_{01,i}, s_{ij}) = \alpha_{01}^{0s_{ij}}(t) e^{\boldsymbol{\beta}_{01}^T \mathbf{z}_{01,ij} + u_{01,i}}$$

$$\alpha_{02}(t|\mathbf{z}_{02,ij}, \boldsymbol{\beta}_{02}, u_{02,i}, s_{ij}) = \alpha_{02}^{0s_{ij}}(t) e^{\boldsymbol{\beta}_{02}^T \mathbf{z}_{02,ij} + u_{02,i}}$$

$$\alpha_{12}(t|\mathbf{z}_{12,ij}, \boldsymbol{\beta}_{12}, u_{12,i}, s_{ij}) = \alpha_{12}^{0s_{ij}}(t) e^{\boldsymbol{\beta}_{12}^T \mathbf{z}_{12,ij} + u_{12,i}}$$

où $\boldsymbol{\beta}_{01}$, $\boldsymbol{\beta}_{02}$, $\boldsymbol{\beta}_{12}$ sont les vecteurs des paramètres de régression, $u_{01,i}$, $u_{02,i}$, $u_{12,i}$ sont les effets aléatoires associés au groupe i et distribués suivant une loi normale tels que :

$$(U_{01}, U_{02}, U_{12})^T \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \mathbf{D} \right).$$

où \mathbf{D} est la matrice de variance-covariance des trois effets aléatoires :

$$\mathbf{D} = \begin{pmatrix} \sigma_{01}^2 & \sigma_{01,02} & \sigma_{01,12} \\ \sigma_{02,01} & \sigma_{02}^2 & \sigma_{02,12} \\ \sigma_{12,01} & \sigma_{12,02} & \sigma_{12}^2 \end{pmatrix}.$$

La vraisemblance du modèle est :

$$\prod_{i=1}^I \frac{\int \int \int \prod_{j=1}^{n_i} \mathcal{L}_{ij}(t_{ij} | u_{01,i}, u_{02,i}, u_{12,i}, s_{ij}) f_{U_{01}, U_{02}, U_{12}}(u_{01,i}, u_{02,i}, u_{12,i}) du_{01,i} du_{02,i} du_{12,i}}{\int \int \prod_{j=1}^{n_i} e^{-A_{01}(t_{0,ij} | \mathbf{z}_{01,ij}, \beta_{01}, u_{01,i}, s_{ij}) - A_{02}(t_{0,ij} | \mathbf{z}_{02,ij}, \beta_{02}, u_{02,i}, s_{ij})} f_{U_{01}, U_{02}}(u_{01,i}, u_{02,i}) du_{01,i} du_{02,i}}$$

où les t_{ij} sont les temps d'évènement et de censure, I est le nombre de groupes, n_i le nombre de sujets du groupe i ; $f_{U_{01}, U_{02}, U_{12}}$ est la fonction de densité de probabilité conjointe des variables U_{01} , U_{02} , U_{12} . On définit les fonctions d'intensité de transition cumulée : $A_{01}(t | \mathbf{z}_{01,ij}, u_{01,i}, s_{ij}) = \int_0^t \alpha_{01}(u | \mathbf{z}_{01,ij}, u_{01,i}, s_{ij}) du$ et $A_{02}(t | \mathbf{z}_{02,ij}, u_{02,i}, s_{ij}) = \int_s^t \alpha_{02}(u | \mathbf{z}_{02,ij}, u_{02,i}, s_{ij}) du$, et $t_{0,ij}$ est l'âge du sujet j du groupe i à l'entrée dans l'étude (dans l'état 0). La troncature à gauche à l'âge d'entrée dans l'étude est représentée par le dénominateur du terme de la vraisemblance.

3 Estimation des paramètres

Afin de pouvoir prendre en compte simplement la troncature et la censure par intervalles nous utilisons soit une approche paramétrique (avec des lois de Weibull) soit une approche par vraisemblance pénalisée. Les paramètres des intensités de transition et de régression sont estimés conjointement en maximisant la vraisemblance du modèle.

Par exemple, la contribution à la vraisemblance d'un sujet entré dans l'étude à l'âge $t_{0,ij}$ et toujours sain en $t_{1,ij}$ est :

$$\mathcal{L}_{ij}(t_{ij} | u_{01,i}, u_{02,i}, u_{12,i}, s_{ij}) = e^{-A_{01}(t_{1,ij} | \mathbf{z}_{01,ij}, \beta_{01}, u_{01,i}, s_{ij}) - A_{02}(t_{1,ij} | \mathbf{z}_{02,ij}, \beta_{02}, u_{02,i}, s_{ij})},$$

avec une contribution au dénominateur de la vraisemblance égale à celle de l'expression présentée au paragraphe précédent :

$$e^{-A_{01}(t_{0,ij} | \mathbf{z}_{01,ij}, \beta_{01}, u_{01,i}, s_{ij}) - A_{02}(t_{0,ij} | \mathbf{z}_{02,ij}, \beta_{02}, u_{02,i}, s_{ij})}.$$

Pour une estimation par maximum de vraisemblance pénalisée les paramètres de lissage sont choisis par validation croisée approchée (Joly et al 2002).

4 Calcul d'espérance de vie

Une fois les intensités de transition estimées elles sont utilisées pour calculer des espérances de vie. Plusieurs espérances de vie différentes peuvent être calculées (Touraine et al, 2013).

Par exemple l'espérance de vie sans démence pour un sujet non dément de 70 ans se calcule ainsi :

$$\frac{\int \int \int_{70}^{+\infty} e^{-A_{01}(t|\mathbf{z}_{01}, \boldsymbol{\beta}_{01}, u_{01}, s) - A_{02}(t|\mathbf{z}_{02}, \boldsymbol{\beta}_{02}, u_{02}, s)} f_{U_{01}, U_{02}}(u_{01}, u_{02}) du_{01} du_{02} dt}{\int \int e^{-A_{01}(70|\mathbf{z}_{01}, \boldsymbol{\beta}_{01}, u_{01}, s) - A_{02}(70|\mathbf{z}_{02}, \boldsymbol{\beta}_{02}, u_{02}, s)} f_{U_{01}, U_{02}}(u_{01}, u_{02}) du_{01} du_{02}}$$

C'est le nombre d'années qu'un sujet vivant et non dément âgé de 70 ans peut espérer encore vivre sans être dément.

Bibliographie

- [1] Duchateau, L. and Janssen, P. (2008), The Frailty Model, Berlin: Springer, Statistics for Biology and health.
- [2] Joly, P., Commenges, D., Helmer, C. and Letenneur, L. (2002), A penalized likelihood approach for an illness-death model with interval-censored data: application to age-specific incidence of dementia, *Biostatistics*, 3, 433-443.
- [3] Joly, P., Gerds, T., Qvist, V., Commenges, D. and Keiding, N. (2012), Estimating survival of dental fillings based on interval-censored data and multi-state models, *Statistics in Medicine*, 31, (11-12):1139-49
- [4] Touraine, C., Helmer, C. and Joly, P. (2013), Predictions in an illness-death model, *Statistical Methods in Medical Research*.