

EVALUATION DU COÛT DE LA MALADIE DE HORTON ET DE SES VARIATIONS À L'AIDE DE MODÈLES DE TYPE POPULATION AVERAGE MODEL UTILISANT LES GENERALISED ESTIMATING EQUATIONS (GEE)

M. Mounié ^{1,2,3} & N. Costa ^{2,3} & G. Pugnet ⁴ & L. Molinier ^{1,2,3} & N. Savy ^{1,5}

¹ *Université Toulouse III, F-31000, Toulouse, France.*

² *INSERM, UMR 1027, F-31000 Toulouse, France.*

³ *Département d'Information Médicale, CHU Toulouse Hôtel Dieu Saint-Jacques, Toulouse F-31059, France.*

⁴ *Service de Médecine Interne, CHU Toulouse Purpan.*

⁵ *Institut de Mathématiques de Toulouse, UMR 5219, CNRS, F-31062, Toulouse, France.*

Résumé. La maladie de Horton est une inflammation de la paroi des artères se localisant principalement au niveau des artères temporales. Elle est ainsi souvent nommée artérite temporale ou artérite à cellules géantes (GCA). De par sa forte incidence et par son grand nombre de symptômes plus ou moins graves, le coût de la maladie de Horton est non négligeable pour l'assurance maladie.

Il est alors nécessaire d'étudier précisément les différents postes de dépenses et leurs survenues afin de pouvoir réfléchir à l'amélioration de sa prise en charge à moindre coût. Le coût de prise en charge des patients atteints de la maladie de Horton est évalué puis discrétisé par catégorie de coût et par tranche de temps (6 mois). Cela permettra de détecter les postes de dépenses les plus importants et leurs survenus. Ces coûts sont ajustés et leurs variations selon différentes covariables sont étudiées à l'aide de modèles de type "population average model" utilisant les "Generalized Estimating Equations" (GEE). Ce type de modèle permet de prendre en compte le caractère longitudinal de nos données en considérant une structure de corrélation inter individus dans l'estimation de nos coefficients. On se concentrera enfin sur les patients atteints de la maladie de Horton en se focalisant sur une complication, la Pseudo Polyarthrite Rhizomélique (PPR) apportant un grand surcoût. Nous avons en notre possession une cohorte de 103 patients atteints de la maladie de Horton et de 606 individus témoins. Après ajustement, la complication PPR entraîne une augmentation moyenne de 60% du coût de prise en charge. De plus, le coût de prise en charge de la maladie est significativement plus important les trois premières années de suivis.

Les résultats obtenus seront comparés aux résultats obtenus à l'aide d'un modèle linéaire généralisé mixte.

Mots-clés. POPULATION AVERAGE MODEL, GENERALIZED ESTIMATING EQUATION, ECONOMIE DE LA SANTE.

Abstract. The Horton disease is an inflammation of the artery walls mainly located in the temporal arteries and so, often called temporal arteritis or Giant Cell Arteritis (GCA). Through its high incidence and its significant number of more or less serious symptoms, the GCA disease cost is significant for health insurance.

So, it becomes necessary to accurately study the various items of expenditures and when they occurred in order to manage and optimize the support of the illness at lower cost.

We assessed the overall average cost first then by category of cost and per time interval of our patients in order to detect the most important drivers cost. Afterwards the cost will be adjust and we appraise their variations according to different covariates with Average Population Models using the Generalized Estimating Equations (GEE). This kind of model allows to take into account longitudinal data by considering an inter individual's correlation structure in the estimation of our coefficients. Finally, we concentrated on Horton's patients by focusing on a complication called polymyalgia rheumatica (PMR) which cost much. In order to do this, we studied a cohort of 103 patients suffering from Horton's disease and 606 control individuals. After adjustment, the PMR complication lead to an 60% increase of the global cost. Beside, he management cost is significantly higher in the first three years.

Obtained result will be compare with a a mixed generalized linear model result.

Keywords. POPULATION AVERAGE MODEL, GENERALIZED ESTIMATING EQUATION, GIANT CELL ARTERISIS, HEALTH ECONOMIC ...

1 Contexte et objectifs

La maladie de Horton est une inflammation de la paroi des artères se localisant principalement au niveau des artères temporales et est ainsi souvent nommée artérite temporale ou artérite à cellules géantes (ACG). Elle est le type le plus commun d'artérite systémique en Europe et Amérique du nord [1]. Elle survient la plupart du temps chez les personnes âgées (>50 ans) avec une incidence de 1/1000 et subit un pic important chez les 75-80 ans avec une incidence de 1/120. [2] Elle est donc amenée à s'accroître parallèlement au vieillissement de la population.

La maladie de Horton entraîne un grand nombre de symptômes tel que des maux de tête, une perte de poids, des douleurs et des difficultés lors de la mastication. De plus, il existe un risque important de troubles visuels allant jusqu'à la perte de la vision [3].

La Pseudo Polyarthrite Rhizomélique (PPR) est un rhumatisme inflammatoire très fréquent après 65 ans (1/2500 [4]) et apparaît chez entre 40% et 60% des patients atteints de la maladie de Horton [5, 6] . Le seul traitement efficace dans la maladie de Horton est une corticothérapie, c'est-à-dire l'administration de cortisone, un anti inflammatoire, pendant une durée minimale de 1 à 2 ans [7]. Le traitement est identique, que le patient

ait ou non une PPR.

De par sa forte incidence et par son grand nombre de symptômes plus ou moins graves, le coût de la maladie de Horton est alors non négligeable pour l'assurance maladie. Il devient alors nécessaire d'étudier précisément le coût global de la maladie de Horton, les différents postes de dépenses et leurs survenus. Pour ce faire, nous allons évaluer le coût direct de la maladie de Horton selon le point de vue de l'assurance maladie et ses variations en se focalisant sur le surcoût apporté par la complication PPR. Aucune étude à ce jour ne traite du coût global de la maladie de Horton et encore moins du surcoût apporté par la PPR.

2 Matériel et méthodes

Nous avons en notre possession, une extraction du Système National d'Information Inter Régime de l'Assurance Maladie (SNIIRAM) relative aux patients Horton en Midi-Pyrénées. Le SNIIRAM nous permet de travailler sur une base médico-administrative exhaustive et en vie réelle pour un horizon temporel important. Nous avons ainsi accès aux informations sur les patients ainsi que toutes les consommations en soins de ville et en hôpital. Les coûts sont discrétisés par catégories de coût et par tranches de temps. Les différentes catégories de coûts sont les hospitalisations, les actes médicaux et les consultations, les actes paramédicaux, les médicaments et les équipements médicaux ainsi que les transports. Ainsi les coûts pris en compte sont les coûts directs médicaux et les coûts directs non médicaux. Nous avons arbitrairement choisi des tranches de temps de 6 mois.

Dans un premier temps, les coûts moyens par tranche de temps et par catégorie ont été comparés afin d'identifier les postes de dépenses les plus importants de la maladie et leurs survenus.

Dans un second temps, le coût moyen de la maladie de Horton est ajusté à l'aide d'un modèle de type "Population Average Model" utilisant les "Generalized Estimating Equation". Ce type de modèle permet de prendre en compte le caractère longitudinal de l'étude en intégrant une structure de corrélation inter-individus dans l'estimation des coefficients [8, 9]. L'écriture du modèle est analogue à celle d'un modèle linéaire généralisé:

$$g(E[Y_{ij}]) = g(\mu_{ij}) = x_{ij}^T \beta \quad (1)$$

pour l'individu i ($i = 1, \dots, K$) au temps j ($j = 1, \dots, n_i$). La fonction g intervenant dans (1) est la fonction de lien et x le vecteur de régression. Soit

$$\text{VAR}(Y_{ij}) = \phi a_{ij} \quad (2)$$

où ϕ est un paramètre de dispersion et $a_{ij} = a(\mu_{ij})$ est une fonction de variance connue. Dans la suite on notera $\vec{Y}_i = (Y_{i,1}, \dots, Y_{i,n_i})$. Soient $\vec{R}_i(\alpha)$, une matrice de corrélation de

travail complètement décrite par un vecteur α à spécifier et la matrice de covariance de travail de \vec{Y}_i :

$$\vec{V}_i = \phi \vec{A}_i^{1/2} \vec{R}_i(\alpha) \vec{A}_i^{1/2} \quad (3)$$

avec \vec{A}_i la matrice diagonale des $(a_{ij}, j = 1 \dots, n_i)$.

Pour les estimations $(\hat{\phi}, \hat{\alpha})$ de (ϕ, α) , l'estimation $\hat{\beta}$ est solution des equations appelées GEE:

$$\sum_{i=1}^K \frac{\partial \vec{\mu}_i^\top}{\partial \beta} \vec{V}_i^{-1} (\vec{Y}_i - \vec{\mu}_i) = \vec{0}. \quad (4)$$

Afin d'estimer les paramètres de la matrice de variance \vec{V}_i et de vérifier l'ajustement du modèle on définit les résidus $e_{ij} = \frac{Y_{ij} - \mu_{ij}}{\sqrt{b''(\theta_{ij})}}$, que l'on évalue en $\beta = \hat{\beta}$. Pour estimer le paramètre de dispersion ϕ , on pose

$$\hat{\phi} = \frac{1}{N - p} \sum_{i=1}^K \sum_{j=1}^{n_i} e_{ij}^2 \quad (5)$$

ou $N = \sum_{i=1}^K n_i$, est le nombre total d'observations dans l'échantillon et $p = \dim(\beta)$

On estime ensuite β en utilisant l'algorithme suivant:

- Estimer β sous l'hypothèse d'indépendance et appeler l'estimateur obtenu $\hat{\beta}_0$.
- Estimer α et ϕ à partir de $\hat{\beta}$ et des e_{ij} .
- Poser $\vec{V}_i = \hat{\phi} \vec{A}_i^{1/2} \vec{R}_i(\hat{\alpha}) \vec{A}_i^{1/2}$
- Mettre la valeur de $\hat{\beta}$ à jour:

$$\hat{\beta}_{m+1} = \hat{\beta}_m + \left(\sum_{i=1}^K \vec{D}'_i \vec{V}_i^{-1} \vec{D}_i \right)^{-1} \left(\sum_{i=1}^K \vec{D}'_i \vec{V}_i^{-1} [\vec{Y}_i - \vec{\mu}_i(\hat{\beta}_m)] \right) \quad (6)$$

Avec $\vec{D}'_i = \frac{\partial \vec{\mu}_i^\top}{\partial \beta}$

- Itérer les étapes 2 à 4 jusqu'à convergence (différence entre $\hat{\beta}_{m+1}$ et $\hat{\beta}_m$ plus petite qu'une tolérance spécifiée).

La matrice de corrélation utilisée étant une matrice de travail, elle est alors possiblement fautive. La variance de $\hat{\beta}$ est alors estimée par un estimateur de matrice de variance sandwich robuste.

Ces modèles considèrent le temps comme une variable explicative et permettent ainsi d'étudier les effets des facteurs et variables globalement au cours du temps. De plus, et contrairement aux modèles mixtes, il n'y a pas de problème quant à la spécification de la distribution jointe des termes d'erreurs et des effets aléatoires. En effet les modèles mixtes cherchent à modéliser la corrélation entre les individus ou groupements tandis que les GEE cherchent plutôt à corriger les estimateurs et les estimateurs de leurs variances et covariances. Un des avantages des GEE sera alors que, malgré une mauvaise spécification de la structure de corrélation, les estimateurs des paramètres du modèle sont convergents. Ils apportent ainsi une meilleure approximation de la réalité [10].

La même méthodologie sera utilisée pour comparer les différences de coûts chez les patients atteints de la maladie de Horton entre les individus atteints de la complication PPR et les autres.

Notre variable réponse sera alors le coût de la maladie de Horton sur des tranches de 6 mois. Notre variable d'intérêt sera si le patient a ou non la complication PPR. De plus, nous avons en notre possession un ensemble de variables décrivant les individus tel que l'âge de l'individu, son sexe, et si il a eu des Antécédents (*Atcd*) de cancer, de diabète, de maladie Cardiovasculaire (*Cvd*), et d'Hypertension (*Hta*). La fonction de lien utilisée sera la fonction log et nos coûts suivent une loi Gamma. La fonction de lien logarithmique est le plus utilisé en économie car étant toujours positive comme nos coûts. La matrice de corrélation utilisée sera une matrice dite auto régressive.

Le modèle complet sera alors de la forme:

$$\begin{aligned} \log(Cout_{ij}) = & GCA_PPR_{ij} + Age_{ij} + Sexe_{ij} + Atcd_Cancer_{ij} \\ & + Atcd_Diabete_{ij} + Atcd_Cvd_{ij} + Atcd_Hta_{ij} + Classe_Temps_{ij} \end{aligned} \quad (7)$$

Enfin, nous utiliserons une démarche pas à pas descendante afin de garder seulement les variables significatives et ayant un effet sur le coût moyen par tranches de temps de la maladie de Horton. Nous garderons néanmoins l'âge et le sexe dans tous les cas.

3 Résultats

L'extraction SNIIRAM est alors composée de 103 patients cas diagnostiqués avec la maladie de Horton entre 2009 et 2015 et de 606 patients témoins appariés avec nos cas par rapport à l'âge et le sexe. A cause d'un suivi trop court, nous enlèverons 7 patients atteints de la maladie de Horton et leurs témoins associés. Nous nous retrouvons alors avec 96 individus atteints de la maladie de Horton et 574 individus témoins. Le coût moyen de prise en charge sur 5 ans des individus atteints de la maladie de Horton est de 28231€ et de 20639€ pour les témoins soit un surcoût apporté par la maladie de 7592€

par patient pour 5 ans dont 6665€ les 3 premières années.

Les résultats obtenus après ajustement par rapport aux antécédents à l'aide du modèle "Population Average Model" restreint aux individus atteints de la maladie de Horton sont consignés dans le tableau 1.

Table 1: Facteurs associés avec la variation du coût moyen des patients atteints de la maladie de Horton sur 6 mois. Atcd: Antécédent

Variable	Modalité	RR	P_Value	IC 95%	
GCA	GCA seul	1			
	GCA _PPR	1.6	0.041	1.02	2.52
Gender	Male	1			
	Female	1.03	0.9	0.67	1.63
Age	<70.6	1			
	[70.6 ; 79.6]	1.17	0.6	0.65	2.09
	>79.6	1.53	0.13	0.88	2.69
Temps	[0 ; 3 an]	1			
]3 ans ; 5 ans]	0.68	0.03	0.51	0.92
Atcd_Diabete	0	1			
	1	2.85	0.0009	1.5	5.31

Après ajustement sur l'âge, le sexe et les antécédents de diabète, un patient atteint de la maladie de Horton avec la complication PPR, aura une prise en charge en moyenne 60% plus coûteuse qu'un individu ayant seulement la maladie de Horton. De plus, il y a une forte baisse du coup de prise en charge de nos patients à partir de 3 ans après le diagnostic.

4 Conclusion

Les modèles de type "Population Average Models" utilisant les GEE, permettent de prendre en compte des données longitudinales et de visualiser l'effet d'un facteur au cours du temps. Ainsi nous avons pu montrer l'augmentation significative du coût de la prise en charge des patients atteints de la maladie d'Horton s'ils ont une PPR. De plus, le coût de la prise en charge devient très important 6 mois après le diagnostic. Ce type de modèle ne demande pas de spécifier de distribution des effets aléatoires et des erreurs. De plus, malgré une mauvaise spécification de la matrice de corrélation, les inférences sur les estimateurs seront encore valides. Une mauvaise spécification de la matrice de corrélation entraînera seulement une variance des estimateurs plus importante. Une des différences principales entre ce type de modèle et les modèles mixtes vient également de

l'interprétation des coefficients. Les modèles mixtes évaluent l'effet d'un facteur sur la variation moyenne de notre variable réponse sachant un individu (ou "cluster") fixé. Les modèles de type "Population Average Models" évaluent l'effet global du facteur sur la variable réponse étudiée, tous individus (ou clusters) confondus. Ainsi, ce type de modèle pourrait être le plus adapté à notre type de données.

References

- [1] Watts RA, Lane SE, Scott DG, Koldingsnes W, Nossent H, Gonzalez-Gay MA, Garcia-Porrúa C, Bentham GA. Epidemiology of vasculitis in Europe. *Ann. Rheum. Dis.* Dec 2001; **60**(12):1156–1157.
- [2] Guillevin L, Pagnoux C. Encyclopédie orphanet grand public 2007. [Www.orpha.net/data/patho/Pub/fr/Horton-FRfrPub876v01.pdf](http://www.orpha.net/data/patho/Pub/fr/Horton-FRfrPub876v01.pdf).
- [3] Gonzalez-Gay MA, Barros S, Lopez-Diaz MJ, Garcia-Porrúa C, Sanchez-Andrade A, Llorca J. Giant cell arteritis: disease patterns of clinical presentation in a series of 240 patients. *Medicine (Baltimore)* Sep 2005; **84**(5):269–276.
- [4] MASSOT C. Pseudopolyarthrite rhizomélique (p.p.r.) 2002. Corpus Médical – Faculté de Médecine de Grenoble.
- [5] Gonzalez-Gay MA. Giant cell arteritis and polymyalgia rheumatica: two different but often overlapping conditions. *Semin. Arthritis Rheum.* Apr 2004; **33**(5):289–293.
- [6] Salvarani C, Cantini F, Boiardi L, Hunder GG. Polymyalgia rheumatica and giant-cell arteritis. *N. Engl. J. Med.* Jul 2002; **347**(4):261–271.
- [7] Fraser JA, Weyand CM, Newman NJ, Biousse V. The treatment of giant cell arteritis. *Rev Neurol Dis* 2008; **5**(3):140–152.
- [8] Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986; **73**(1):13–22, doi:10.1093/biomet/73.1.13. URL <http://dx.doi.org/10.1093/biomet/73.1.13>.
- [9] Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* Dec 1988; **44**(4):1049–1060.
- [10] Hubbard A, Ahern J, Fleischer N, Van der Laan M, Lippman S, Jewell N, Bruckner T, WA S. To gee or not to gee : Comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. *Lippincott Williams Wilkins* 2010; .