International Journal of Innovation and Scientific Research
ISSN 2351-8014 Vol. 6 No. 1 Aug. 2014, pp. 56-62

http://www.ijisr.issr-journals.org/

Modélisation des Facteurs de Risque Génétiques dans l'Accident Vasculaire Cérébral Ischémique

[Modeling of Genetic Risk Factors in Ischemic Stroke]

Khalid BALAR, Sellama NADIFI, Khalil HAMZI, and Bréhima DIAKITE

Laboratory of Human Genetics and Molecular Pathology,
University Hassan II/ Faculty of Medicine,
Casablanca, Morocco

ABSTRACT: In this article, we focus our work on the modeling of genetic risk factors on ischemic strokes occurred. To do this, logistic regression was widespread in our study. We proceeded in two stages: the first, we modeled the probability of the occurrence of ischemic stroke in an individual (i) based on genetic risk factors. Our sample consisted of 330 individuals aged at least 40 years, divided into 165 patients who had an ischemic stroke and 165 controls.
We applied the Wald test for all variables in the model one by one and we concluded to Reject H0, since the coefficients of our variables are not all zero.
In a second step, we studied the effects of these variables on the risk factors and then the effect of variables on Ischemic stroke to present the model equation.
We set a prediction threshold after specification test, that allows us to ensure the quality of the fit of the model and its degree of prediction, the proportion of people who have ischemic stroke is (50%). The results showed that 128 of 156 people with Ischemic stroke allowed a positive predictive value of 82%. We conclude that the prediction rate and the success rate of our model is 80.30 %, the results obtained with the «XLSTAT» software show a very good model with (sensitivity 78% and specificity 83%).

KEYWORDS: Ischemic stroke, genetic risk factors, modeling, logistic regression and fit test.

RÉSUMÉ: Dans cet article, nous focalisons notre travail sur la modélisation des facteurs de risque génétiques sur la survenu des AVC ischémiques. Pour ce faire, La régression logistique a été largement répandue dans notre étude. Nous avons procédé en deux étapes, dans la première, nous avons modélisé la probabilité de la survenue d'un accident vasculaire cérébral ischémique chez un individu (i) en fonction des FR génétiques. Notre échantillon était composé de 330 individus âgés d'au moins 40 ans, répartis en 165 patients ayant fait un AVCI et 165 témoins.
Nous avons appliqué le test de Wald, pour toutes les variables du modèle une à une et nous avons conclu au Rejet de H0, puisque les coefficients de nos variables ne sont pas tous nuls.
Dans une seconde étape, nous avons étudié les effets de ces variables sur les FR et ensuite l'effet des variables sur l'AVCI afin de présenter l'équation du modèle.
On a fixé un seuil de prédiction après test de spécification qui nous permet de nous assurer de la qualité de l'ajustement du modèle et son degré de prédiction, la part des personnes qui ont AVCI (50%). Les résultats obtenus ont montré que 128 personnes sur 156 ayant un AVCI ont permis une valeur de prédiction positive de 82%. Nous concluons que le taux de prédiction ou le taux de succès de notre modèle est de 80.30%, les résultats obtenus avec le logiciel «XLSTAT» montrent un très bon modèle (sensibilité de 78% et spécificité de 83%).

MOTS-CLEFS: AVC ischémique, facteurs de risques génétiques, modélisation, régression logistique et Test d'ajustement.

Corresponding Author: Khalid BALAR

1 INTRODUCTION

L'accident vasculaire cérébral ischémique (AVCI) est une pathologie fréquente, grave et invalidante, reconnue comme problème majeur de santé publique.

Elle représente la première cause de handicap non traumatique acquis chez l'adulte et la deuxième cause de démence après la maladie d'Alzheimer et la troisième cause de mortalité. [1]

L'incidence de cette pathologie augmente nettement avec l'âge, or la population mondiale vieillit : entre 2009 et 2050 le nombre de personnes âgées de plus de 60 ans va tripler dans le monde, et celui des plus de 80 ans va quadrupler. Ceci laisse présager une forte croissance de la prévalence de l'AVCI au cours du siècle. [2], [3].

Selon la définition de l'Organisation mondiale pour la santé, l'AVCI est un déficit neurologique focal (ou parfois global) d'apparition soudaine, durant plus de 24 heures d'origine vasculaire [4].

Le Maroc est sérieusement menacé par les maladies cardiovasculaires qui constituent un enjeu épidémiologique. Parmi celles-ci, on trouve les accidents vasculaires cérébraux ischémiques qui constituent actuellement un véritable problème de santé au Maroc.

Une enquête épidémiologique réalisée à Rabat et Casablanca, en 2010, avec le soutien financier de l'Académie Hassan II des sciences et techniques, a montré que la prévalence des AVC au Maroc était de 300 pour 100.000 habitants. Cette étude a été faite sur 30 000 ménages en milieux urbain et rural Cette enquête, constitue la première du genre en Afrique et dans le monde arabe, elle a regroupé - plusieurs universitaires marocaines, (neurologues, généticiens, biologistes, statisticiens, cardiologues, endocrinologues et nutritionnistes) . Cette étude avait pour objectif d'évaluer le rôle des affections cardiaques, des facteurs nutritionnels, biologiques et génétiques dans la genèse de l'accident vasculaire cérébral ischémique (AVCI) [5].

Le but de ce travail est de Modéliser l'impact des FR génétiques sur la survenue d'un AVCI.

2 METHODOLOGIE

2.1 MATERIEL

Il s'agit d'une étude Cas Témoins apparié réalisée pendant 3 années dans le laboratoire de génétique et pathologie moléculaire de la faculté de médecine de Casablanca (LGPM) en collaboration avec les services de neurologie des CHU de Casablanca et Rabat.

Notre échantillon de départ était composé de plus 725 individus dont 195 ont un AVCI. Nous avons retiré de cet échantillon les individus de moins de 40 ans, L'échantillon restant comporte 330 individus âgés de 40 ans et plus, dont 165 sont malades, et 165 témoins apparié sur l'âge chaque cas d'AVCI à 1 témoin. Puisque d'après la littérature[7], l'âge est le facteur de risque le plus important, en effet après 55 ans, et pour chaque tranche d'âge de 10 ans, les taux d'AVCI sont multipliés par 2 à la fois chez l'homme et chez la femme[8],[9].

Le LGM nous a confié les résultats de l'analyse génétique des 330 échantillons, cette analyse a concerné 9 gènes (MTHFR, FII, ACE, FV, APOE, PAI, ENOS, APOA5, ALOX5AP).

2.2 METHODE

Les données de notre échantillon sont analysées à l'aide du logiciel de statistique le plus complet de Microsoft «XLSTAT» qui est basée sur le langage Visual Basic. Le code de XLSTAT utilise à la fois ce code (VBA, pour l'affichage) et du code C++ (pour les calculs), compatible avec les plateformes Windows et Mac.

Nous avons procédé à la réalisation d'un modèle de régression logistique [10],[11],[12]. Ce dernier permet de prédire la probabilité pour qu'un AVCI arrive (valeur de 1) ou non (valeur de 0) à partir de l'optimisation des coefficients de régression.

Le résultat varie toujours entre 0 et 1. [13] Lorsque la valeur prédite est supérieure à 0,5, l'événement (AVCI) est susceptible de se produire, alors que lorsque cette valeur est inférieure à 0,5, il ne l'est pas.

Le modèle de régression logistique a comporté plusieurs étapes :

1- une recherche bibliographique approfondie au préalable est obligatoire. En effet la qualité d'une régression logistique repose, avant tout, sur le choix des variables explicatives [15], [16] que l'on est susceptible d'intégrer au modèle.

2- Il a été nécessaire ensuite d'étudier et d'analyser les liaisons entre chacune des variables explicatives [cf. tableau I] et la variable dépendante: on a réalisé une analyse univariée ; les odds-ratios calculés sont bruts

3-Nous avons été contraints d'essayer plusieurs stratégies afin de parvenir à un modèle final qui devrait porter le maximum d'informations tout en ayant un nombre limité de variables, afin de faciliter l'interprétation : les plus employées sont les procédures dites « pas à pas descendantes ou pas à pas ascendantes ».

- **Présentation du Modèle**

Il s'agit de modéliser la probabilité de la survenue d'une maladie cardiovasculaire chez un individu i en fonction des facteurs génétiques.

AVCI = 1 s'il y a AVCI
AVCI = 0 si pas d'AVCI

Ou AVCI est une variable latente qui peut s'écrire comme la somme d'¦une combinaison linéaire de caractéristiques propres à chaque individu et d'un terme aléatoire.

AVC= β xi + ε i

- x est un vecteur de variables explicatives ;
- β est le vecteur associe des paramètres ;
- ε est l'aléa ;

Afin de calculer la probabilité, il a été nécessaire de spécifier une distribution statistique pour les εi. Les deux lois statistiques les plus couramment utilisées sont la loi logistique et la loi normale, qui donnent alors le modèle qualitatif binaires appelé Logit « Le modèle Logit offre un avantage sur le plan de la technique d'estimation des paramètres et son fondement mathématique est relativement simple » [14].

Le modèle Logit suppose que F suit une loi logistique. Dans ces conditions, la probabilité qu'un individu ait une AVCI s'écrit :

$$P(AVCI=1) = \frac{exp^{(\beta i X i)}}{1+exp^{(\beta i X i)}}$$

Par conséquent, la probabilité de ne pas avoir cette maladie sera donnée par :

$$P(AVCI=0) = \frac{1}{1+exp^{(\beta i X i)}}$$

- **Ecriture du modèle**

AVCI = β0+ β1âge + β2FII + β3 MTHFR + β4APOE + β5F5C + β6PAI + β7APOA5 + β8ALOX5AP

Tableau 1. Décrit les variables explicatives utilisées dans notre modèle

Variable	Modalités	Effectifs	%
FII	AA	5	1,52
	GA	49	14,85
	GG	276	83,64
MTHFR	CC	149	45,15
	CT	151	45,76
	TT	30	9,09
ACE	DD	188	56,97
	ID	115	34,85
	II	27	8,18
APOE	E2/E2	4	1,21
	E2/E3	30	9,09
	E2/E4	47	14,24
	E3/E3	193	58,48
	E3/E4	33	10,00
	E4/E4	23	6,97
F5C	CC	176	53,33
	CT	112	33,94
	TT	42	12,73
ENOS	GG	190	57,58
	GT	117	35,45
	TT	23	6,97
PAI	4G	122	36,97
	4G/5G	90	27,27
	5G	118	35,76
APOA5	CC	27	8,18
	CT	100	30,30
	TT	203	61,52
Alox5AP	AA	181	54,85
	TA	93	28,18
	TT	56	16,97

3 RESULTATS

L'échantillon comporte 330 cas témoins apparié selon l'âge avec une moyenne d'âge de 55 ans

Les résultats sont présentés sous forme de moyennes, les données sont analysées à l'aide du logiciel «XLSTAT». Le test de signification entre chaque facteur de risque avec l'AVCI à partir du test Ki Deux

La régression logistique et le coefficient de corrélation ont été utilisés pour analyser les relations existant entre l'AVCI, l'âge, le sexe, FII, MTHFR, APOE, F5C, ACE, PAI, APOA5, ALOX5AP.

Nous avons choisis comme tests d'hypothèses, le test sur les paramètres (test de Wald) et le test de spécification (tableau de contingence).

Le test de Wald, proche du test de score, sert spécifiquement à tester la nullité d'un ou plusieurs coefficients, en particuliers de tous sauf la constante.

H0= tous les coefficients sont nuls / H1= au moins un des coefficients est différent de 0.

On a appliqué le test de Wald, pour toutes les variables du modèle une à une et on a conclu au Rejet de H0, puisque les coefficients de nos variables ne sont pas tous nuls. [cf. tableau 2].

Tableau 2. Test de l'hypothèse nulle H0 : Y=0,500 (Variable AVCI)

Statistique	DDL	Khi²	Pr > Khi²
-2Log(Vraisemblance)	22	190,799739	< 0,0001
Score	22	144,511677	< 0,0001
Wald	22	81,9472572	< 0,0001

Une fois que le modèle de prédiction, a été conçu, nous avons évaluer l'efficacité et l'ajustement. On a pu le faire de la manière suivante :

Confronter les valeurs observées de la variable dépendante avec les prédictions [cf. tableau 3].

En utilisant le Test de spécification [17] qui nous a permis de nous assurer de la qualité de l'ajustement du modèle et de son degré de prédiction et de calculer par la suite le pourcentage d'observations bien prédites qui donne un critère de performance du modèle.

Tableau 3. Classification pour l'échantillon d'estimation (Variable AVCI)

	PREDICTION		
OBS	0	1	Total
0	**137**	28	165
1	37	**128**	165
Total	174	156	330

On a fixé un seuil de prédiction, la part des personnes qui ont AVC (165/ 300=0,5) dont on a eu comme résultats, 128 personnes qui ont un AVC ont été bien prédis sur 156, avec une valeur de prédiction positive de 82,05% et 137 qui n'ont pas AVC ont été bien prédis sur 174. Le taux de prédiction de notre modèle est de 80,30% ((128+137/330*100). [cf. tableau 4]

Tableau 4. Indicateurs de notre Modèle

Vrais positifs	128
Faux positifs	37
Taux d'erreur	19,70%
Taux de succès	**80,30%**
Sensibilité	77,58%
Spécificité	83,03%

Les résultats obtenus avec le logiciel «XLSTAT» montrent un très bon modèle (sensibilité de 77,58%% et spécificité de 83%).

La courbe ROC, évaluant les résultats de classification en fonction du seuil de décision est sensibilité en fonction de spécificité (Figure 1). Cette courbe ROC montre, l'AUC surface sous la courbe ROC (Area Under Curve) égale 0,89 (ce qui conduit à une bonne sensibilité) comme nous l'avions souligné, et plus l'AUC est grand, meilleur est le test.

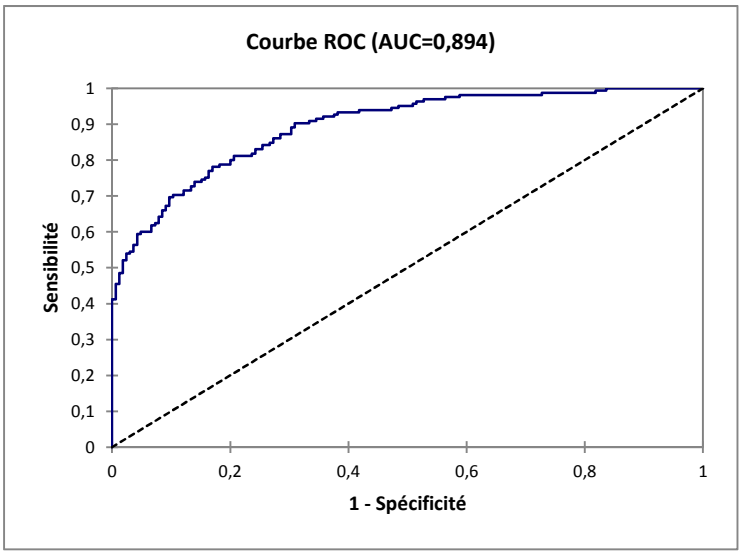

Fig. 1. Courbe ROC du seuil de décision, pour une modélisation de FRG sur l'AVCI.

4 CONCLUSION

Les résultats de modélisation ont montré que les facteurs génétiques étaient des facteurs de risque puissant pour les AVC ischémiques

On a observé que le modèle de la régression logistique dans notre étude cas Témoins, nous a permis d'analyser la corrélation entre la survenue d'un AVCI et de ses facteurs génétiques.

L'outil informatique et ses applications nous ont permis de réaliser plus aisément cette analyse. Cependant, dans la matrice de confusion, nous avons conclu, le modèle de prédiction réalise 28 + 37 = 65 mauvaises prédictions. Le taux d'erreur est de 65/330 = 19,7 %

La statistique du rapport de vraisemblance LAMBDA est égale à 190, la probabilité critique associée est 0. Le modèle est donc globalement très significatif, il existe bien une relation entre les variables explicatives (FII, MTHFR, APOE, F5C, ACE, PAI, ENOS, APOA5, ALOX5AP…) et la variable expliquée.

Apres une étude individuelle, des coefficients liés à chaque variable explicative, nous avons constaté que les gènes ACE et ENOS ne semblent pas jouer de rôle significatif dans cette analyse.

REFERENCES

[1] Murray, Lopez et al. Motality by cause for eight regions in the world : Global Burden of Disease Study, Lancet,1997, 349, pp. 1269-1276

[2] Murray, Lopez et al., Alternative projections of mortality and disability by cause 1990- 2020: Global Burden of Disease Study, Lancet, 1997, 349, pp. 1498-1504

[3] « World population to exceed 9 billions by 2050 », World Population Prospects: The 2008 Revision – Press Release – March 2009, disponible sur http://www.un.org/esa/population/publications/wpp2008/pressrelease.pdf Accédé le 04/10/2009

[4] WHO STEP Stroke Manual : The WHO STEPwise approach to Stroke Surveillance – World Health Organisation, 2005

[5] A.Cherkaoui. Accident vasculaire cérébral ischémique: une étude de grande ampleur au Maroc

[6] ROSENBERG N, MURATA M, IKEDA Y. The frequent 5, 10-methylenetetrahydrofolate reductase C677T polymorphism is associated with a common haplotype in whites, Japanese an Africans. Am J Hum Genet 2002; 70: 758-62.

[7] Di Carlo, Lamassa et al., Stroke in the very old: clinical presentation and determinants of 3-month functional outcome: A European perspective. European BIOMED Study of Stroke Care Group, Stroke, 1999, 30, pp. 2313-2319

[8] Nakayama, Jorgensen et al., The influence of age on stroke outcome: the Copenhague Stroke Study, Stroke, 1994, 25, pp. 808-813

[9] Noone, O'Shea et al., Stroke in the very old, Ir Med J., 2008, 101 (1), pp. 8-9

[10] Aminot I, Damon MN The Use of Logistic Regression in the Analysis of Data Concerning Good Medical Practice

[11] J. Jaccard, Intercation E_ects in Logistic Regression, Series: Quantitative Applications in the Social Sciences, n0135, Sage Publications, 2001.
[12] D. Garson, Logistic Regression, http://www2.chass.ncsu.edu/garson/PA765/logistic.htm
[13] P.L. Gonzales, "Modèles à réponses dichotomiques", in Modèles statistiques pour données qualitatives, Droesbeke, Lejeune et Saporta Editeurs, Chapitre 6, pages 99-136, Technip, 2005.
[14] R. Rakotomalala, Régression logistique - Une approche pour rendre calculable P (Y/X), http://eric. univ-lyon2.fr/~ricco/cours/supports_data_mining.html
[15] S. Menard, Applied Logistic Regression Analysis (Second Edition), Series: Quantitative Applications in the Social Sciences, n0106, Sage Publications, 2002.
[16] A.A. O'Connell, Logistic Regression Models for Ordinal Response Variables, Series: Quantitative Applications in the Social Sciences, n0146, Sage Publications, 2006.
[17] D.W. Hosmer, S. Lemeshow, Applied Logistic Regression, Second Edition, Wiley, 2000.