



Prévision de la séroprévalence de l'hépatite A dans la wilaya de setif en 2024

FORECAST OF HEPATITIS A SEROPREVALENCE IN THE WILAYA OF SETIF IN 2024

Bourioune Tahar¹, GUENIFI Wahiba²

¹ Faculté des sciences économiques, gestion et sciences commerciales, Université FERHAT Abbas Setif 1- Algérie 19000

² Faculté de médecine, Université FERHAT Abbas Setif 1- Service des maladies infectieuses, CHU de Sétif, Algérie 19000

Résumé

La séroprévalence des anticorps anti hépatite A (correspondant au taux d'immunisation) était de 100% à Sétif (Algérie), chez les personnes âgées entre 10 et 14 ans en 1986. Elle est passée à 70,4% en 2011. Partant de ce fait, les auteurs se proposent de prévoir cette séroprévalence, dans la même wilaya en 2024, à travers le modèle 'Logit binaire multiple', sur la base des données d'une enquête réalisée en 2011. La séroprévalence globale chez les sujets âgés entre 5 et 19 ans serait, selon les résultats de ce modèle, de 67% en 2024 ; les principaux facteurs associés à cette séroprévalence seraient l'âge, l'habitat, la taille des ménages et l'antécédent d'ictère. En conséquence, un programme de vaccination pourrait s'imposer comme une nouvelle stratégie de lutte contre la maladie dans la wilaya de Sétif.

Mots clés : Hépatite A, Séroprévalence, Prévision 2024, Facteurs de risque, Sétif.

Abstract

The seroprevalence of anti-hepatitis A antibodies (corresponding to the immunization rate) was 100% in Sétif, in people aged between 10 and 14 years in 1986. It has declined to 70.4% in 2011. Starting from this fact, the authors propose to predict this seroprevalence, in the same wilaya (district) in 2024, through the 'multiple binary logit' model, based on data from a survey carried out in 2011. The overall seroprevalence in subjects aged between 5 and 19 years would be, according to the results of this model, 67% in 2024; the main factors associated with this seroprevalence would be age, habitat, household size and a history of jaundice. As a result, a vaccination program could establish itself as a new disease control strategy in Sétif.

Keywords: Hepatitis A, Seroprevalence, Forecast 2024, Risk factors, Sétif.

Correspondance :

Guenifi Wahiba et al., Faculté de médecine, Université FERHAT Abbas Setif 1- Service des maladies infectieuses, CHU de Sétif, Algérie 19000, **Téléphone** : +213667339852- **Email** : wahibaguenifi@yahoo.fr

Article reçu : 10-11-2022

Accepté : 05-12-2022

Publié : 25-01-2023



Copyright © 2023. GUENIFI WAHIBA et al.

This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Pour citer cet article : Guenifi Wahiba et al. Prévision de la séroprévalence de l'hépatite A dans la wilaya de setif en 2024. Revue de Médecine et de Santé Publique. 2023 ; 6(1) : 126 - 137.

1- Introduction

L'amélioration du niveau d'hygiène a considérablement modifié le profil épidémiologique de l'hépatite virale A (HVA) dans de nombreuses régions du monde au cours des dernières décennies. Son incidence a beaucoup diminué chez l'enfant et la maladie a glissé vers l'adulte ; par conséquent plus de cas symptomatiques et de formes graves sont notifiés et le risque de survenue d'épidémie par accumulation de personnes non immunisées devient plus important, incitant ainsi à revoir, dans les pays de faible endémicité, les modalités préventives et la mise en place d'un programme de vaccination [1]. À Sétif, il a été rapporté, sur une étude menée en 2011, la décroissance de la séroprévalence des anticorps antiviral de l'hépatite A (anti-VHA) dans la wilaya de Sétif et par suite le passage de cette dernière de la zone de haute endémicité vers la zone d'endémicité moyenne. À cette date, la séroprévalence globale (entre 5 et 19 ans) rapportée était de 72,28 %. Elle était étroitement liée à l'âge : le taux est passé de 56,9 % chez les enfants âgés de 5 - 9 ans à 70,4 % chez ceux âgés entre 10 et 14 ans et atteint 85,4 % entre 15 et 19 ans ; par contre en 1986, 100 % des personnes prélevées, âgées de 10 à 14 ans, avaient des anticorps anti-VHA [2]. Selon la littérature empirique, ce type de changement épidémiologique est signalé depuis plus d'une décennie dans d'autres pays de la région du Moyen-Orient et de l'Afrique du Nord ; par exemple, en Tunisie, en 2008, une étude nationale, retrouve une séroprévalence de 83,9 % chez les enfants de 5 à 9 ans contre 99 % avant l'âge de 5 ans pour les années 1980 [3]. En Egypte, le taux de séroprévalence rapporté en 2008 était de 56 % avant l'âge de 10 ans et de 73,8 % entre 10 et 18 ans [4] alors que durant les années 1990, la séroprévalence avant l'âge de 10 ans était de 100 % [5]. Dans les pays développés, les modifications épidémiologiques signalées ont entraîné plus précocement le passage progressif de certains pays à un niveau d'endémicité modéré voire faible [6].

L'objet de ce travail est un essai de modélisation de la séroprévalence des anticorps anti-VHA dans la wilaya de Sétif, au moyen du modèle Logit binaire multiple, sur la base des résultats d'un questionnaire effectué en 2011, dans le but de prévoir la séroprévalence des anticorps anti-VHA dans cette wilaya du pays en 2024. Méthodologiquement, pour l'étude du modèle sus cité, nous utiliserons les techniques d'analyse statistique (descriptive et inférentielle) et de prédiction des modèles économétriques à variables qualitatives [7].

2- Matériel et méthodes

2.1- Données

Les données sont issues de l'enquête effectuée en 2011 sur une population cible de 226 433 sujets, âgés de 5 à 19 ans, de la wilaya de Sétif [2]. La population d'étude, issue d'un échantillonnage aléatoire stratifié a considéré 1061 sujets répartis selon un choix raisonné sur 4 dairates de cette Wilaya (Sétif, Ain oulmène, Ain arnat et Ain kebira). 15 variables pouvant être associées à la séroprévalence des anticorps anti-VHA ont été étudiées dans cette enquête (en plus du sexe et de l'âge, il y avait 4 variables socio-économiques, 3 variables cliniques, 3 variables concernant l'alimentation en eau et 3 variables appréciant le niveau d'hygiène) ; ce sont des variables explicatives de type quantitatif ou qualitatif. La variable anti-VHA représente la variable dépendante ; elle est de type nominale dichotomique (Négatif /Positif).

La variable statistique « âge » a été classifiée selon la méthode heuristique en 3 classes : '5-9ans', '10-14ans' et '15-19 ans'. Le nombre de personnes vivant sous le même toit correspondant à la variable « taille des ménages » a été réduit à 3 catégories : réduite (2-4 personnes), moyenne (5-8 personnes) et large (≥ 9 personnes). Le niveau d'éducation des parents (chaque parent considéré à part), variable qualitative ordinale à 5 catégories, a été réduit à 2 catégories : inférieur (regroupant les niveaux illettré, primaire et moyen) et supérieur (regroupant les niveaux secondaire et universitaire).

La variable « habitat » a 2 catégories : rural et urbain et la variable « source d'alimentation » en eau est également à 2 catégories (puits/forage et réseau d'alimentation en eau potable).

Les autres variables sont des variables qualitatives dichotomiques à 2 catégories : oui et non ; il s'agit des variables : antécédent d'ictère, antécédent d'ictère dans l'entourage, antécédent d'HVA dans l'entourage, consommation d'eau minérale en bouteille, restauration à l'extérieur, lavage des crudités et des fruits, lavage des mains avant les repas et lavage des mains après être allée aux toilettes.

2.2- Modèles théoriques

Un modèle à variable qualitative a été choisi parmi l'ensemble des modèles économétriques [8,9].

2.2.1. Formalisation

Parce que les modèles à variables qualitatives sont appropriés au type de notre variable expliquée ; nous avons opté pour le modèle « Logit binaire multiple ». Ce dernier est un modèle statistique, à l'instar du modèle « Probit », où la variable dépendante est qualitative nominale binaire et les variables indépendantes sont un mélange de variables quantitatives et catégorielles [10,11,12].

Soit y_i une variable nominale binaire $y_i \in \{0,1\}$ / $y_i \in \{0,1\}$, $P(y_i=1)=p_i$; $P(y_i=0)=1-p_i$.

Essayons de regresser y_i sur k prédicteurs $(x_{i1}, x_{i2}, \dots, x_{ik}) = X_i'$ à l'instar de la régression linéaire

$y_i = X_i' \beta + u_i$. Comme il est impossible de définir la régression à partir de $p(y=+/-x) = a_0 + a_1 x_1 + \dots + a_k x_k$, le premier terme étant dans $\{0,1\}$ alors que le second est dans \mathbb{R}^k ;

une solution à ce problème passerait par la création d'une variable latente y^* telle que:

$y = 1 \Leftrightarrow y^* = X_i' \beta + u_i > 0$. Il s'ensuit : $p_i = P(y_i = 1) = P(y^* > 0) = P(X_i' \beta + u_i > 0) = P(u_i > -X_i' \beta) = 1 - F(-X_i' \beta)$ c.a.d. $y_i \in \{0,1\}$ avec $p_i = 1 - F(-X_i' \beta) = \frac{1}{1 + e^{-X_i' \beta}}$

$F(X_i' \beta) = \frac{1}{1 + e^{-X_i' \beta}}$ pour le Logit d'où $X_i' \beta = \ln\left(\frac{p}{1-p}\right) = \ln(\text{Odds } p)$.

$F(X_i' \beta) = \Phi(X_i' \beta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{X_i' \beta} e^{-z^2/2} dz$ pour le Probit d'où $X_i' \beta = \Phi^{-1}$.

2.2.2. Estimation des paramètres

i – Calcul du gradient $\nabla_{\beta} L(\beta)$

$$\begin{aligned} \nabla_{\beta} L(\beta) &= \nabla_{\beta} \left[\sum_i y_i x_i' \beta - \ln(1+e^{x_i' \beta}) \right] = \left[\sum_i \nabla_{\beta} y_i x_i' \beta - \ln(1+e^{x_i' \beta}) \right] = \left[\sum_i \nabla_{\beta} (y_i x_i' \beta) - \nabla_{\beta} [\ln(1+e^{x_i' \beta})] \right] \\ &= \sum_i y_i x_i - x_i \frac{e^{x_i' \beta}}{1+e^{x_i' \beta}} = \sum_i x_i (y_i - p_i) = \mathbf{X}^T (\mathbf{Y} - \mathbf{Y}') \quad / \mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nk} \end{pmatrix}_{n,k} : \text{"design matrix "}; \end{aligned}$$

$(y_i - y_i')_{n,1}$

ii – Calcul du Hessian $\nabla_{\beta\beta} L(\beta)$

$$\begin{aligned} \nabla_{\beta\beta} L(\beta) &= \nabla_{\beta} \left[\sum_i x_i (y_i - p_i) \right] = \sum_i \nabla_{\beta} [x_i (y_i - p_i)] = \sum_i \nabla_{\beta} [-p_i x_i] = \sum_i \nabla_{\beta} \left[-\frac{1}{1+e^{-x_i' \beta}} x_i \right] = \\ &= \sum_i -(-1) \left[\frac{1}{1+e^{-x_i' \beta}} \right]^2 (-x_i) e^{-x_i' \beta} x_i = - \sum_i \left[\frac{1}{1+e^{-x_i' \beta}} \right] \left[\frac{e^{-x_i' \beta}}{1+e^{-x_i' \beta}} \right] (-x_i) x_i = - \mathbf{X}^T \mathbf{W} \mathbf{X} \\ \mathbf{W} &= \begin{pmatrix} p_1(1-p_1) & 0 \\ 0 & p_n(1-p_n) \end{pmatrix}; \end{aligned}$$

iii- Calcul des paramètres optimaux β^* par la méthode itérative de Newton-Raphson

$$\beta_{t+1} = \beta_t - H^{-1} g$$

$$\beta_{t+1} = \beta_t - \frac{\nabla_{\beta} L(\beta)}{\nabla_{\beta\beta} L(\beta)} \Leftrightarrow \beta_{t+1} = \beta_t + (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{Y} - \mathbf{Y}')$$

Remarque : si H n'est pas inversible, il faut appliquer Levenberg Marquardt method ou gradient descent

3- Résultats

Nos résultats regroupent les statistiques descriptives, les tests non paramétriques d'association et la régression logistique binaire univariée puis multiple [12,13,14].

Au préalable, nous avons codifié les variables qualitatives car concernant les facteurs de risque qualitatifs, l'analyse économétrique exige la codification de leurs catégories et la désignation parmi ces dernières, d'une catégorie de référence.

Les variables à 3 catégories (âge et taille des ménages) ont été codifiées (0), (1) et (2) telles que la tranche d'âge '5-9 ans' et la taille des ménages 'large' ont le code (0). Pour les variables dichotomiques codifiées en (0) et (1), le code (1) a été attribué à la catégorie 'urbain' pour la variable habitat, au 'réseau d'alimentation en eau potable' pour la source d'alimentation en eau,

au niveau 'supérieur' pour l'instruction des parents et au sexe 'masculin'. Concernant les variables qualitatives dichotomiques à catégories oui et non, le code (1) a été attribué à la catégorie 'oui'. Le code (0) représente la catégorie de référence dans notre 'étude.

Les premières analyses entreprises sur l'ensemble des variables originelles (régression logistique univariée et tests statistiques non paramétriques d'association) nous ont conduits à retenir 9 variables sur les 15 variables originelles. Le tableau 1 résume les résultats des tests caractéristiques des variables retenues.

Tableau I : Variables significatives associées à la séroprévalence en régression univariée et tests statistiques non paramétriques d'association

Variabes	χ^2	p_value	Wald	p_value
Age par tranches	70.14	0.000	65.70	0.000
Habitat	11.44	0.001	11.18	0.001
taille des ménages	12.81	0.002	12.51	0.002
Niveau d'éducation du père	4.72	0.03	5.047	0.025
Niveau d'éducation de mère	5.40	0.02	5.766	0.016
Antécédents d'ictère	10.62	0.001	10.05	0.002
Antécédent d'ictère dans l'entourage	5.74	0.017	6.098	0.014
Antécédent d'HVA dans l'entourage	5.30	0.021	5.543	0.019
Consommation d'eau minérale en bouteille	4.86	0.028	5.453	0.02

3-1. Modèle logit optimal

Par suite, une première analyse de régression logistique multiple entreprise sur les 9 variables retenues, avec le logiciel SPSS18, a conduit, après avoir dissipé les effets de confusion inter-variables, à sélectionner 4 variables indépendantes significatives à savoir l'âge, l'habitat, la taille des ménages et l'antécédent d'ictère. Elles constituent le modèle Logit I de notre étude (Tableau 2).

Tableau II: Estimation du modèle LOGIT I

	B	Erreur std	Wald	Degrés de liberté	Signif .	Exp(B)	IC à 95% pour Exp(B)	
							Borne inférieure	Borne supérieure
Age par tranche 05-9 ans			69,169	2	,000			
Age par tranche (1)10-14 ans	,667	,171	15,297	1	,000	1,948	1,395	2,722
Age par tranche (2)15-19 ans	1,582	,190	69,162	1	,000	4,866	3,351	7,064
Habitat (1)	-,816	,202	16,372	1	,000	,442	,298	,656
Taille des ménages > 9p			6,863	2	,032			
Taille des ménages (1) 5-8p	-,360	,210	2,928	1	,087	,698	,462	1,054
Taille des ménages (2) 2-4p	-,739	,282	6,850	1	,009	,478	,275	,831
Antécédent d'ictère (1) 8	1,008	,290	12,112	1	,001	2,740	1,553	4,834
Constante	1,150	,256	20,164	1	,000	3,157		

Ces 4 variables sont toutes significatives à un seuil de confiance de 95%. Le tableau 3 présente la codification de ces facteurs.

Tableau III : Codification des variables catégorielles

			Fréquence	Paramètres de Codification	
				(1)	(2)
Age par tranche	Age par tranche (0) :05-9ans	par	297	,000	,000
	Age par tranche (1) :10-14ans	par	361	1,000	,000
	Age par tranche (2) :15-19ans	par	403	,000	1,000
	Taille des m		109	,000	1,000

Taille des ménages	Taille des ménages (2)	771 181	1,000 ,000	,000 ,000
Antécédent d'ictère	Non (0) Oui (1)	951 110	,000 1,000	
Habitat	Non (0) Oui (1)	845 216	1,000 ,000	

Il reste à déterminer les termes d'interactions – au cas où elles existent - entre les différents variables. Après prospection, il s'avère qu'il existe une interaction significative entre les facteurs de risque « âge » et « habitat ». Le modèle Logit II ci-après présente alors le modèle Logit avec l'inter-action « âge par tranche (0)*habitat (1) ».

Tableau IV : Estimation modèle Logit II

	B	Erreur std	Wald	df	Sig.	Exp(B)	IC à 95% pour Exp(B)	
							Borne inférieure	Borne supérieure
Age par tranche 05-9ans			6,011	2	,050			
Age par tranche(1) 10-14ans	-,358	,397	,813	1	,367	,699	,321	1,523
Age par tranche(2) 15-19 ans	,904	,512	3,123	1	,077	2,470	,906	6,731
Habitat (1)	-1,421	,309	21,12	1	,000	,242	,132	,443
Taille des ménages > 9p			7,548	2	,023			
Taille des ménages (1) 5-8p	-,379	,212	3,204	1	,073	,685	,452	1,037
Taille des ménages (2) 2-4p	-,784	,286	7,534	1	,006	,456	,261	,799
Antécédent d'ictère (1)	1,058	,292	13,17	1	,000	2,882	1,627	5,104
Age par tranche*Habitat			8,302	2	,016			
Age par tranche (1) – Habitat (1)	1,255	,440	8,129	1	,004	3,508	1,480	8,312
Age par tranche (2) – Habitat (1)	,835	,551	2,299	1	,129	2,305	,783	6,783
Constante	1,636	,327	25,08	1	,000	5,137		

Les résultats de ce tableau affirment l'augmentation de la séropositivité avec l'âge et l'antécédent d'ictère et sa diminution avec l'amélioration des conditions d'habitat et la réduction de la taille des ménages.

Par exemple, la probabilité d'être séropositif pour la tranche d'âge 15-19 ans est de 2,47 fois supérieure rapporté à la tranche d'âge 5-9 ans. Le surplus relatif de séropositivité pour un antécédent d'ictère est de 2,88. La probabilité d'être séropositif pour l'habitat urbain est de 0,758 inférieure par rapport à l'habitat « rural ».

Le facteur d'interaction « âge-habitat » est significatif au seuil de confiance de 95% : degré de signification=0.016. La probabilité d'un individu qui appartient à la tranche d'âge « 10-14ans » et dont le mode d'habitation est 'urbain' a un surplus de probabilité d'être séropositif de 3,51 fois relativement à l'enfant de « 5-9 ans » qui habite la campagne.

La probabilité d'un individu qui appartient à la tranche d'âge « 15-19ans » et dont le mode d'habitation est 'urbain' a un surplus de probabilité d'être séropositif de 2,30 fois relativement à l'enfant de « 5-9 ans » qui habite la campagne.

3.2 . Prédiction

Le modèle prédictif se formalise ainsi :

$$\hat{Y} = \ln(\hat{p}/(1-\hat{p})) = 1,636 - 0,358 * \text{âge par tranche (1)} + 0,904 * \text{âge par tranche (2)} - 1,421 * \text{habitat(1)} - 0,379 * \text{taille des ménages(1)} - 0,784 * \text{tailles des ménages(2)} + 1,058 * \text{antécédent d'ictère(1)} + 1,255 * (\text{âge par tranche(1)} * \text{habitat(1)}) + 0,835 * (\text{âge par tranche(2)} * \text{habitat(1)}).$$

3.3. Prévision

Sur la base des statistiques des prévisions démographiques [14], si l'on suppose qu'en 2024, dans la wilaya de Sétif, le mode d'habitation urbain deviendra prédominant [habitat(1)] et la taille des ménages passera à la taille réduite (2 à 4 personnes) [taille des ménages(2)]. En se basant sur les trois autres facteurs prédictifs (âge, antécédent d'ictère et âge*habitat), les prévisions obtenues par le modèle Logit II sont présentées dans le tableau suivant :

Tableau V : Résultats des prévisions par le modèle Logit II

X_i	$\text{Log}(p/1-p)$	$(p/1-p)$	p	$Y=\text{indicateur}$
0,1,2,0,0	-0,569	0,566	0,361	0
0,1,2,1,0	0,489	1,631	0,620	1
1,1,2,0,1	0,328	1,388	0,581	1
1,1,2,1,1	1,744	5,720	0,851	1
2,1,2,0,0	-0,569	0,566	0,361	0
2,1,2,1,0	0,489	1,631	0,620	1

$X'_i = (\text{âge}, \text{habitat(1)}, \text{taille ménages(2)}, \text{antécédent d'ictère}, \text{âge*habitat})$

Quand pour un individu donné, \hat{p} est $> 0,5$ alors cet individu a des anticorps anti VHA (Y=positif) c.à.d. l'indicateur Y=1.

La séroprévalence générale prévisionnelle en 2024 par le modèle II équivaldrait à 67%.

5. Discussion

Le modèle Logit II est apprécié sur la base du pseudo coefficient de détermination R^2 , du test d'adéquation aux données de Hosmer & Lemeshow et de la matrice de confusion [15].

Le pseudo R^2 de Nagelkerke vaut 0,155; ceci signifie que le modèle Logit II explique seulement 15,50% de la 'variance totale'. L'hypothèse nulle du 'bon ajustement du modèle aux données' de Hosmer et Lemeshow est acceptée (degré de signification = 98,70%). Par conséquent, le modèle peut être considéré satisfaisant (Tableaux 6 et 7).

Tableau VI : Résumé du modèle

	-2 Log likelihood	Cox & Snell R Square	Nagelkerke Square	R
1	1132,216 ^a	,107	,155	

Pseudo R^2 indique dans quelle mesure le modèle prédit l'output Y. Il se calcule et s'interprète de façon analogue au R^2 de la régression linéaire de Pearson. Pseudo $R^2 \approx 1$ indique une bonne prédiction du modèle; Pseudo $R^2 \approx 0$, indique une mauvaise prédiction du modèle [16,17,18,19].

$$R^2 \text{ de PEARSON(1896)} = 1 - \frac{SCR}{SCT} \quad \text{scr=somme carrés des résidus; sct=somme carrés totale}$$

$$\text{Pseudo } R^2 \text{ de Mc Fadden(1974)} = 1 - \frac{LL_{\text{modèle}}}{LL_{\text{null}}} \quad LL_{\text{modèle}}=\text{déviance résiduelle; } LL_{\text{null}}=\text{déviance nulle}$$

$$\text{Pseudo } R^2 \text{ de Cox \& Snell(1989)} = 1 - \left(\frac{LL_{\text{null}}}{LL_{\text{modèle}}} \right)^{2/N}$$

$$\text{Pseudo } R^2 \text{ de Nagelkerke(1991)} = \frac{R^2 \text{ Cox \& Snell}}{1 - LL_{\text{null}}^{2/N}}$$

Tableau VII : Test de Hosmer et Lemeshow

Step	Chi-square	df	Sig.
1	,973	6	,987

Le test Hosmer Lemeshow (LM) indique la qualité d'ajustement du modèle aux observations. C'est un test de chi-carré [20] :

$$HL = \sum_{i=1}^g \sum_{j=1}^k \frac{(obs_{ij} - exp_{ij})^2}{exp_{ij}}$$

obs_{ij} : fréquences observées

exp_{ij} : fréquences prévues

g = Le nombre de groupes.

$g - k$ = degrees of freedom;

H_0 : le modèle ajuste bien les données, différence non significative entre les fréquences observées et prévues.

H_1 : Le modèle n'ajuste pas bien les données ; différence significative entre les fréquences observées et prédites.

En outre, la matrice de confusion (Tableau 8) indique que le pourcentage des prédictions correctes, relativement à la distribution effective, est de 73,40%. Ce qui consolide notre jugement antérieur.

Tableau VIII : Table de classification^a: Matrice de Confusion

Observé			Prédit		
			Anti VHA	positif	Pourcentage correcte
Step 1	Anti VHA	Négatif (0)	85	209	28,9
		Positif (1)	73	694	90,5
Pourcentage global					73,4

a. la valeur seuil est .50

En résumé, le modèle II est acceptable bien qu'il puisse être amélioré par l'adjonction d'autres variables significatives.

Finalement, le modèle 'Logit', qui présuppose une distribution logistique non évidente, s'adapte mieux aux études épidémiologiques comportant d'innombrables variables catégorielles dont l'interprétation mathématique de leurs outputs est facilitée par les ratios d'ODDS et d'OR. Le modèle 'Probit', son principal concurrent, présuppose une distribution normale, non évidente non plus, d'interprétation mathématique lourde, il s'adapte mieux à la toxicologie.

5. Conclusion

Ce travail a consisté à construire un modèle de prédiction de l'infection par le VHA afin de prévoir le taux d'immunisation pour l'année 2024 à Sétif. Ce modèle a été jugé économétriquement acceptable, bien qu'il soit perfectible par l'ajout d'autres facteurs de risque potentiels. La séroprévalence prévisible par ce modèle Logit II est une fonction croissante de l'âge et de l'antécédent d'ictère. Par contre, elle décroît en fonction de l'amélioration des conditions d'habitat et de la diminution de la taille des ménages. La séroprévalence globale a été estimée en moyenne à 67% pour l'année 2024. Le taux des personnes immunisés va diminuer et par conséquent un programme de vaccination devrait être discuté afin de minimiser le risque d'épidémie et de formes sévères. En perspective, Il nous reste à améliorer la capacité prédictive du modèle Logit II en lui incorporant des variables significatives complémentaires.

Référence

1. LARRUE H., ABRAVANEL F., PERON J.-M. Hépatite A et sa vaccination. *EMC – Hépatologie* 2020 ; 7-015-B-22. https://www-clinicalkey-fr.snd11.arn.dz/service/content/pdf/watermarked/51-s2.0-S1155197620597866.pdf?locale=fr_FR&searchIndex=
2. GUENIFI W., LAOUAMRI S., LACHEHEB A. Changes in prevalence of hepatitis A and associated factors in Setif-Algeria. *Revue d'Epidémiologie et de Santé Publique*, 65, 437-442, 2017. <https://doi.org/10.1016/j.respe.2017.05.009>
3. REZIG D., OUNEISSA R., MHIRI L., MEJRI S., HADDAD-BOUBAKER S., BENALAYA N., TRIKI H. Seroprevalences of hepatitis A and E infections in Tunisia ». *Pathol Biol (Paris)*, 56(3):148-53, May 2008. doi: 10.1016/j.patbio.2007.09.026. Epub 2008 Feb 21.
4. AL-AZIZ A.M., AWAD M.A. Seroprevalence of hepatitis A virus antibodies among a sample of Egyptian children. *East Mediterr Health J*, 14: 1028-35, 2008. <https://pubmed.ncbi.nlm.nih.gov/19161074/>
5. DARWISH M.A., FARIS R., CLEMENS J.D., RAO M.R., EDELMAN R. High seroprevalence of hepatitis A, B, C, and E viruses in residents in an Egyptian village in The Nile Delta: a pilot study, *Am J Trop Med Hyg*, 54: 554-8, 1996. doi: 10.4269/ajtmh.1996.54.554.
6. JOUSSEMET M., DEPAQUIT J., NICAND E., MAC NAB C., MEYNARD J.B., TEYSSOU R., FABRE G., BUISSON Y. Effondrement de la séroprévalence de l'hépatite virale A chez les jeunes Français. *Gastroentérol Clin et Biol*, 23(4):447-52, Mai 1999. doi : GCB-05-1999-23-4-0399-8320-101019-ART98
7. ROTHMAN K.J., Greenland S. and LASH T.L. *Modern epidemiology*. , 2008, ed. Lippincott Williams & Wilkins. ISBN: 0781755646, 9780781755641
8. BADI HANI BATTAGI. *Econometrics*. 2011, ed. Springer Verlag,, pages 331-358. <https://doi.org/10.1007/978-3-642-20059-5>
9. GUJARATI D. *Econometrics by example*. 2011, ed. Palgrav-Macmillon, 142-155. ISBN-10: 0230394353
10. O'CONNEL A. *A Logistic regression models for ordinal response variables. Series : quantitative applications in social sciences*, Sage Publications, International Educational AND Professional Publisher Thousand Oaks, London, 2006.
11. CARRIERE I. Comparaisons des méthodes d'analyse des données binaires ou ordinales corrélées. Application à l'étude longitudinale de l'incapacité des personnes âgées. *SCIENCES DU VIVANT [Q-*

- BIO]. Université Paris Sud, PARIS, 2005. Français. <https://tel.archives-ouvertes.fr/tel-00107384/document>.
12. HURLIN C. Econometrie des variables qualitatives ; les modèles multinomiaux. *Ordonnées et non Ordonnés Cours ESA*, Université D'Orleans, France, 2003. https://www.univ-orleans.fr/deg/masters/ESA/CH/Qualitatif_Chapitre2.pdf
13. HOSMER D.W., LEMESHOW J.S., STURDIVANT R.X. *Applied logistic regression, third edition*, WILEY SERIES IN PROBABILITY AND STATISTICS SERIES, 2013, 35. <https://www.wiley.com/en-us/Applied+Logistic+Regression%2C+3rd+Edition-p-9780470582473>
14. *Rapport national sur les objectifs du Millénaire pour le développement (Algérie)*. Edité par le gouvernement algérien, Juillet 2005, ISBN : 9947-0-0961-0
15. RAKOTOMALALA R. *Pratique de la Régression Logistique Régression Logistique Binaire et Polytomique*, Université Lumière Lyon 2, 13-May-2017/8:21. https://eric.univ-lyon2.fr/~ricco/cours/cours/pratique_regression_logistique.pdf
16. MCFADDEN, D. Conditional logit analysis of qualitative choice behavior, P. Zarembka (Ed) *Frontiers in econometrics*. 1974, 104-142, New York: Academic Press. <https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf>
17. COX, D. R., & SNELL, E. J. *The analysis of binary data*, 2ème ed. London: Chapman and Hall, 1989. ISBN 9780412306204.
18. NAGELKERKE, N. A note on a general definition of the coefficient of determination, *Biometrika*, 1991, 78, 691–692. <https://doi.org/10.1093/biomet/78.3.691>
19. THOMAS J. SMITH et CORNELIUS M. MCKENNA. A Comparison of Logistic Regression Pseudo R² Indices, *Multiple Linear Regression Viewpoints*, 2013, Vol. 39(2). https://www.glmj.org/archives/articles/Smith_v39n2.pdf
20. HOSMER D.W., LEMESHOW S. A goodness –of –fit test for the multiple logistic regression model. *Communications in Statistics A10*, 1980, 1043-1069.