

VALIDITÉ DES FACTEURS DE RISQUE
IDENTIFIÉS AU VIH/SIDA
Étude de cas sur le Cameroun

Rapport de recherche

Juin 2007

Nom de l'étudiant : Georges BOADE

Directeur de recherche : Claude MONTMARQUETTE

Sommaire

Depuis 1981, date au cours de laquelle les premiers cas furent identifiés aux États-Unis, la pandémie du VIH a de nos jours en son actif des millions de décès. Cette pandémie est particulière dans ce sens qu'elle décime surtout les populations sexuellement active, et donc économiquement productives. Ceci a un impact non seulement sur le niveau de la production globale, mais encore il faut des ressources supplémentaires pour la prise en charge des malades et des enfants orphelins issus de couples victimes de la contamination.

Sur le plan épidémiologique, des études sont faites, et des facteurs d'exposition sont identifiés comme expliquant la prévalence du VIH.

L'existence des rapports descriptifs ne suffit pas à conclure de la pertinence de ceux-ci. Dans ce sens, à travers ce rapport de recherche, nous posons la question de la « validité des facteurs de risque identifiés au VIH/SIDA ».

Par le biais d'un modèle logistique, nous passons le test de significativité de certains facteurs redondants dans la littérature du VIH/SIDA, et dont les statistiques existent dans notre base de données. Ces données sont issues de la troisième enquête démographique et de santé au Cameroun, effectuée en 2004 et concernant un échantillon de 12062 individus (6038 hommes et 6027 femmes) d'âge compris entre 15 et 59 ans pour les hommes, et de 15 à 49 ans pour les femmes.

Nous constatons que si certains le sont même avec des seuils trop petits, d'autres ont un effet explicatif négligeable. Ainsi l'âge, le lieu de résidence, le sexe, le niveau d'éducation et la taille du ménage sont à retenir dans les politiques d'ensemble. La religion et le statut matrimonial sont à intégrer dans l'approche genre car, elles influencent dans la sérologie des femmes.

En marge de ces résultats, il ressort aussi que ces modèles ne se substituent pas au test du VIH en laboratoire, et ne peuvent être utilisés pour discriminer entre personnes saines et porteuses.

Table des matières

Sommaire.....	1
Liste des tableaux.....	4
Abréviations.....	5
Introduction.....	6
1- Revue de la littérature.....	8
a) Cartographie de l'infection au VIH.....	8
b) Présentation du Cameroun.....	8
c) Études antérieures.....	10
Modèles d'incubation.....	10
La transmission.....	11
Risque et Estimation de la population infectée.....	12
2- Analyse théorique et descriptif des données.....	15
Approche théorique.....	15
Continuité de l'étude.....	15
Les données.....	16
Échantillonnage.....	16
Le test en Laboratoire.....	17
3- Modèle de validation.....	19
a) Analyse descriptive.....	19
Le niveau d'éducation (EDUCAT).....	19
Le lieu de résidence (RESIDENC).....	19
Le statut matrimonial (STATMATR).....	20
La taille du ménage (TAILLEME).....	20
La religion (RELIGION).....	20
Statut à l'emploi (EMPLOI).....	20
Fumeur (FUMEUR).....	21
Sexe (SEX), âge (AGEIND) et âge au premier rapport sexuel.....	21
b)- Analyse multi variée.....	21
Effet marginal.....	23
Prédiction des probabilités.....	24
Effet marginal sur la probabilité d'un évènement.....	24
c)- Implications sur les politiques publiques.....	25
Conclusion.....	26
Bibliographie.....	27
Annexes.....	28
Annexe 1 : Prévalence par tranche d'âge et par sexe.....	29
Annexe 2 : Résultats d'ensemble.....	30
Tableau 2.1: Résumé des observations prises en compte dans le modèle.....	30
Tableau 2.2 : classification en catégories.....	30
Tableau 2.3 : Codification automatique des variables (regresseurs).....	31
Annexe 3: Resultats pour hommes.....	32
Tableau 3.1: Résumé des observations prises en compte dans le modèle.....	32
Tableau 3.2 : classification en catégories.....	32
Tableau 3.3 : estimés des coefficients des regresseurs.....	33
Annexe 4: Resultats pour femmes.....	34
Tableau 4.1: Résumé des observations prises en compte dans le modèle.....	34
Tableau 4.2 : Classification en catégories.....	34
Tableau 4.3 : estimés des coefficients des regresseurs.....	35

Liste des tableaux

Tableau 1.1	Séroprévalence	Page 29
Tableau 2.1	Résumé des observations prises en compte dans le modèle d'ensemble	Page 30
Tableau 2.2	classification en catégories	Page 30
Tableau 2.3	Codification automatique des variables (regresseurs)	Page 31
Tableau 3.1	Résumé des observations prises en compte dans le modèle (Hommes)	Page 32
Tableau 3.2	classification en catégories	Page 32
Tableau 3.3	Estimés des coefficients des regresseurs	Page 33
Tableau 4.1	Résumé des observations prises en compte dans le modèle (femmes)	Page 34
Tableau 4.2	Classification en catégories	Page 34
Tableau 4.3	Estimés des coefficients des regresseurs	Page 35

Abréviations

VIH:	Virus de l'Immuno Déficience Humaine
SIDA:	Syndrome d'Immuno Déficience Acquise
WHO:	World Health Organization
ONUSIDA:	Programme Commun des Nations Unies pour le VIH/SIDA
UNAIDS:	The joint United Nations programme on HIV/AIDS
USAIDS:	United States AIDS programme

Introduction

La pandémie du VIH est source de nombreux défis pour la communauté internationale.

Sur le plan médical, il n'existe pas de vaccin contre le VIH, de même qu'il n'existe pas de traitement pour sauver les personnes infectées du spectre de la mort.

Sur le plan politique, la proportion de personnes infectées est inconnue, et seuls des proportions estimées sont à la base des programmes d'intervention. Les tests de dépistages sont faits sur une base volontaire.

Les médicaments pour soulager les patients restent encore dans la plupart des cas non accessibles, en particulier dans les pays en développement où l'État doit aussi faire face à la paupérisation des populations.

La transmission du VIH se fait soit verticalement, c'est-à-dire de la mère à l'enfant, ou horizontalement, donc entre une personne infectée et une autre qui est saine suivant d'autres formes.

Les études visant à identifier les conditions de contamination révèlent de nombreux cofacteurs, ainsi que des populations dites à haut risque. Ces voies de contamination épousent les réalités socio-économiques du milieu, car les modes de contamination sont fonction des méthodes sexuelles pratiquées, et dépendent donc des zones géographiques.

En général les sous groupes à hauts risque se recrutent dans les populations démunies, et celles se livrant à des activités impliquant l'échange d'instruments ayant été en contact frais avec le sang d'une autre personne. Par exemple, nous pouvons citer les personnes qui utilisent des drogues injectables, les travailleuses de nuit, les enfants de rues, etc...

En Afrique par exemple, l'hétérosexualité est la pratique la plus courante. C'est pour cette raison que les estimations des personnes infectées se font via les femmes enceintes, et constituent de bons estimés (WHO, 2000).

La prise en compte des facteurs favorisant la transmission du virus est indispensable pour la planification des interventions visant à réduire la prévalence. Par conséquent une allocation optimale des ressources nécessite que des facteurs significatifs soient identifiés. C'est dans le but de contribuer à la sélection, parmi de nombreux candidats que nous avons choisi de questionner la validité des facteurs déjà identifiés.

Dans notre travail, nous prenons pour acquis les cofacteurs existants, et essayons de construire un test de passage basé sur le modèle logistique. Les données sont issues de l'enquête démographique et de santé, collectées au Cameroun en 2004.

La variable à expliquer est du type dichotomique et prend les valeurs 0 si la personne n'est pas infectée, et 1 sinon.

Toutefois, les résultats obtenus sont générés par un test en laboratoire qui ne permet pas d'identifier les personnes fraîchement contaminées, car il est basé sur la détection des anticorps dirigés contre le virus du VIH. Ceci signifie donc que les porteurs sains seraient à priori dans la classe nulle. Dans cette situation, afin de tenir compte de cet artefact, les résultats inférés intègrent des coefficients de correction. Mais, pour notre travail, nous n'utilisons que les résultats du test tel que révélé en laboratoire.

Notre travail est reparti en trois chapitres. Nous utilisons à la fois une analyse descriptive et une approche multi variée : l'analyse descriptive est faite en croisant les résultats du test avec les cofacteurs candidats, et l'analyse multi variée tient compte de tous à la fois.

Le premier chapitre est focalisé sur la revue de littérature ainsi que sur l'ampleur de l'infection au VIH à l'échelle mondiale. L'approche théorique et une analyse descriptive des données à notre disposition sont abordées dans le deuxième chapitre, alors que le troisième est concentré sur le schéma de validation des cofacteurs. Ainsi il existe un modèle d'analyse multi variée d'ensemble, un modèle pour les hommes et aussi un pour les femmes tous sont basés sur les mêmes variables afin de faciliter les comparaisons. Pour cela, la classe d'âge de plus de 49 ans qui est particulière aux hommes a été exclue à cette étape.

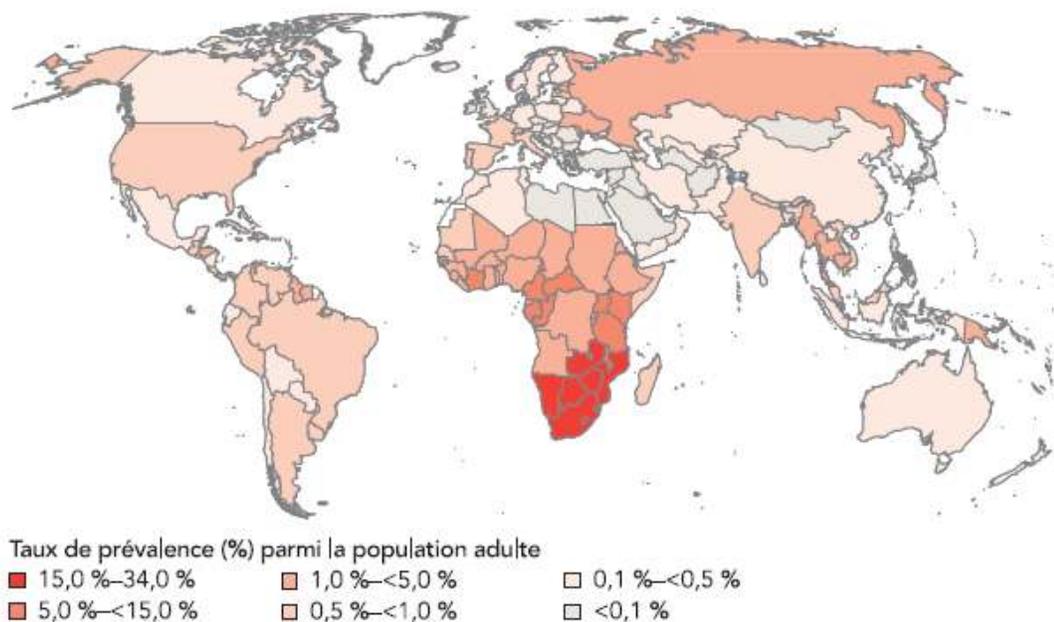
Chacun des trois modèles fait l'objet d'une analyse, et enfin une conclusion en est tirée, celle-ci met en exergue l'importance de notre démarche dans les politiques publics en matière de lutte contre le VIH au Cameroun.

1- Revue de la littérature

a) Cartographie de l'infection au VIH

L'ONUSIDA estime à 38.6 millions de personnes vivant avec le VIH en 2006, contre 33,4 millions en 2005. Les statistiques collectées par l'ONUSIDA montrent une prévalence plus visible en Afrique au sud du Sahara, et une prévalence atténuée sur le reste du globe.

Figure 1.1 : prévalence dans le monde



Source : ONUSIDA, 2006

En Afrique subsaharienne, le taux de prévalence est passé de 6.2% en 2003 contre 6.1% en 2005. 60% de la population mondiale infectée s'y trouve, alors que sa part dans la population du globe n'est que de 10%. Bien que stable, on a noté des progrès dans les pays comme le Burundi, la Côte d'Ivoire et le Kenya. Par contre certains pays sont en nette régression, comme l'Afrique du Sud et le Lesotho (ONUSIDA 2006).

L'infection de personnes par le VIH entraîne une prise en charge de celles-ci, et une perte de ressources humaines dans la suite.

b) Présentation du Cameroun

Le Cameroun est un pays de forme triangulaire situé en Afrique centrale et abritant plus de 16 millions d'âmes. D'une superficie de 475442 km², le Cameroun partage une frontière

commune avec le Nigeria, le Tchad, la République Centrafricaine, le Gabon, le Congo Brazzaville et la Guinée Équatoriale.

Sa situation géographique et le dynamisme de sa population font de lui un pays stratégique dans la sous région Afrique Centrale.

Sa diversité culturelle (plus de 280 langues locales auxquelles s'ajoutent deux langues officielles que sont l'anglais et le français), son climat et son sol permettent une diversification des productions agricoles en toute saison, faisant ainsi de lui le grenier agricole de la sous région. Ces caractéristiques lui valent le pseudonyme d' «Afrique en miniature ».

Le Cameroun jouit d'une stabilité politique, et le régime est semi présidentiel. La vie politique est animée par près d'une centaine de parties politiques, mais une dizaine de ceux-ci sont présents en permanence sur le terrain politique.

L'économie repose essentiellement sur l'agriculture d'exportation. A la fin des années 1980s, le Cameroun est frappé par une crise économique, de laquelle il se remet peu à peu.

Le taux de prévalence au VIH est passé de 0.5% en 1987 à 6.5% en 1998. De nos jours il est stabilisé au tour de 5.5%. Sa croissance notoire jusqu'en 1998 est à concilier avec la période de crise économique.

Les données sur la prévalence que nous utilisons concernent les hommes de 15 à 59 ans, et les femmes de 15 à 49 ans exclusivement.

Les cofacteurs pouvant expliquer la prévalence sont de plusieurs ordres : le niveau d'éducation- le milieu de vie (rural ou urbain)- le sexe- l'âge- la classe sociale- le statut matrimonial- le statut par rapport à l'emploi- le nombre de partenaires sexuels- la circoncision, etc.

Le test du VIH est très coûteux, c'est pour cela que dans la pratique, les tests ELISA et Western-blot qui sont plus abordables sont pratiqués. Ils ne détectent pas le virus, mais plutôt la présence d'anticorps dirigés contre le VIH.

Donc, malgré leur fiabilité, ces tests ne détectent pas les individus récemment contaminés, car la période d'incubation va de quelques semaines à deux ans en moyenne selon les individus, pour que l'organisme secrète les anticorps (Comby 1989).

c) Études antérieures

Il existe des modèles mathématiques de certains aspects de la maladie, en particulier la phase d'incubation a fait l'objet de plusieurs approches.

Modèles d'incubation

La période d'incubation ou séroconversion, qui est l'intervalle de temps séparant l'instant de l'infection à celui où l'on manifeste les premiers signes, est de deux(2) mois.

Il existe des modèles mathématiques de distribution des périodes d'incubation.

Soit U la variable aléatoire représentant la période d'incubation, alors $F(t) = P(U < t)$ est la probabilité pour un individu infecté d'évoluer vers la maladie en t années.

La fonction de survie $S(t) = 1 - F(t)$ et la densité de probabilité de la période d'incubation est $dF/dt = F'(t) = f(t)$.

La fonction de hasard $h(t) = f(t)/S(t)$ est le risque de développer la maladie après t périodes sachant que l'individu n'a pas été infecté avant t périodes.

Il existe des méthodes d'estimation paramétriques et non paramétriques, mais les modèles paramétriques sont appropriés en épidémiologie du HIV (Cox & Oakes 1984).

Dans la pratique, le hasard pour le HIV est connu pour être trop petit juste après l'infection, puis croissant mais incertain après 6 ans du fait que les patients sont pris en charge (Brookmeyer & Mitchell 1994).

Compte tenu de cette description, les distributions pouvant décrire l'incubation sont entre autres :

a) La distribution de Weibull

$$F(t) = 1 - \exp(-\lambda t^p) \text{ avec } p, \lambda > 0$$

La fonction de hasard $h(t) = \lambda p t^{p-1}$ est croissante si $p > 1$, décroissante si $p < 1$ et constante si $p = 1$.

Ce modèle, avec $p > 1$, a été utilisé par Brookmeyer & Goedert (1989)

Darby(1990), Lui, Darrow & Rutherford (1988), Lawrence & Morgan (1986), Medley, Anderson, Cox (1987).

Ce modèle présente des limites à long terme, car son hasard est indéfiniment croissant

b) La loi Gamma

La densité de probabilité de la loi Gamma est

$\lambda^k t^{k-1} \exp(-\lambda t) / \Gamma(k)$ $\lambda, k > 0$ et $\Gamma(\cdot)$ est la fonction gamma.

Ce modèle renseigne lorsque λ est supposé constant et $k > 1$. Il a été utilisé dans les études de l'incubation chez les personnes infectées lors de transfusions sanguines (Medley, Anderson, Cox, et Billard 1987).

c) Loi log-logistic

$F(t) = 1 - (1 + (\lambda t)^\beta)^{-1}$ avec $\lambda, \beta > 0$

Pour $\beta > 1$, le hasard croît d'abord, jusqu'à atteindre un maximum, puis décroît par la suite lorsque t tend vers l'infini. Il a été utilisé par Lui et al (1988) pour modéliser l'incubation chez les homosexuels hommes aux États unis.

d) Les modèles mixtes

On suppose qu'il existe une proportion α de personnes infectées de période d'incubation distribuée suivant F_1 , et $1-\alpha$ distribuée suivant F_2 . alors $F(t) = \alpha F_1(t) + (1-\alpha)F_2(t)$

Ce modèle a été utilisé par Rutherford et al. (1988), Degruittolla et al. (1988)

La transmission

Le VIH se transmet par voies sanguines et sexuelles. Cette transmission peut se faire verticalement, par exemple de la mère à l'enfant, ou horizontalement c'est à dire entre deux individus à l'exemple de partenaires sexuels, de l'utilisation des mêmes instruments comme les drogués qui échangent les seringues.

Les cofacteurs de risques relèvent donc du comportement des individus, et forment deux groupes (Flaskerud & Ungvarski 1994) :

➤ Les cofacteurs d'exposition sont ceux qui augmentent la probabilité de contacter le virus. Exemple de la sexualité anale passive, les partenaires sexuels multiples, la présence d'ulcères génitaux, le partage d'aiguilles et seringues, l'utilisation des drogues récréatives, la transfusion sanguine, un membre de l'entourage porteur du virus.

➤ Les cofacteurs de déclenchement accélèrent l'immunodéficience et aggravent le risque d'être infecté pendant l'exposition, ou contribuent au déclenchement d'une maladie active chez les porteurs.

Les cofacteurs de déclenchements non infectieux sont entre autres la malnutrition, l'usage des drogues prescrites, les facteurs génétiques, le stress émotionnel, l'âge, la grossesse, le sexe.

Parmi les cofacteurs de déclenchements infectieux, on peut citer le surmenage antigénique lié à des infections multiples, les infections virales et immunosuppression (Hépatite B, herpès, etc.)

Risque et Estimation de la population infectée

a) Risque et facteurs de risques

Si nous supposons deux cohortes 1 et 2, et qui sont suivies de la période 0 à la période t, le risque relatif permet de comparer le niveau de risque entre les deux cohortes. Il est donné par $rr(t) = k_2(t) / k_1(t)$ où $k_i(t) = h_i(t) / (1 - H_i(t))$ est le taux de hasard de la cohorte i, $h_i(t) = dH(t) / dt$ est la fonction de densité et $H_i(t)$ la fonction cumulative de la période d'incubation.

Le risque est associé généralement avec la fréquence des activités sexuelles, les maladies sexuellement transmissibles, l'usage des drogues, le nombre de partenaires sexuels dans l'actif, les voyages dans les zones infectées.

Padran et al. (1991) ont rapporté que le taux de transmission qui se fait de l'homme infecté à la femme saine est plus élevé que celui de la femme à l'homme sain. Dans leur étude, ils rapportent 1.4% contre 20%, et confirmé par une étude européenne (De Vincenzi 1992), qui rapporte 12% contre 20%. La comparaison de ces deux rapports suggèrerait-elle que les cofacteurs facilitant la transmission dépendraient de la zone géographique?

b) Estimation de la population

La maîtrise de la taille des populations infectées qui passe par l'identification des sous-groupes à risque est capitale pour plusieurs raisons, (WHO 2002):

- Politique publique
 - Besoin de planification des interventions publiques
 - une allocation efficace des ressources
 - estimations du nombre de personnes infectées et projections
- Besoins de programmation
 - planification des interventions
 - mesure de la couverture
 - le suivi et l'évaluation des programmes d'intervention

Mais, cette taille de personnes infectées reste un défi à relever par la communauté internationale.

Les méthodes d'estimation varient d'un pays à l'autre, en l'absence des enquêtes statistiquement viables.

En général les mesures de surveillance sont centrées sur les sous populations ou le taux de prévalence est le plus concentré. Les groupes à risque les plus reconnus dans la littérature sont : les hétéro et homosexuels, les utilisateurs de drogues injectables, les personnes utilisant des produits sanguins.

En Afrique subsaharienne par exemple les estimations sont faites sur la base des femmes enceintes qui vont passer leurs visites prénatales à l'hôpital. On part du fait que la contamination dans ces pays se fait surtout par voie de rapports hétérosexuels, et donc les femmes infectées constitueraient une bonne base d'inférence sur la population totale (WHO). Il est important de relever ici que les tests effectués sur l'échantillon de sang prélevé sont un acte de routine.

Il existe aussi des campagnes de dépistage gratuit et des tests de routine faits sur des malades qui vont se faire consulter dans les centres médicaux.

Sur le plan statistique, ces échantillons ne sont pas représentatifs de la population totale du pays, et donc les estimations sont sujettes à un biais de sélection.

En Amérique Latine, en Europe de l'Est ou en Asie par exemple, cette approche serait inappropriée. Chaque groupe à risque ayant ses cofacteurs.

En effet, la zone géographique et les comportements propres des personnes doivent être pris en compte dans l'identification des sous populations à hauts risques.

Les comportements risqués peuvent changer au cours des ans, s'adaptant aux réalités politiques et socio-économiques (WHO 2004).

Par exemple, dans certaines villes et campagnes, la présence des prostituées obéit à un cycle d'affaire local au cours de l'année. Donc, une estimation en leur présence sans en tenir compte lors de la planification aura un impact sur l'efficacité des interventions et projections. Ainsi, les facteurs suivants sont à prendre en compte dans la dynamique des interventions : la mode, la conjoncture économique, les réseaux de distribution des drogues, les migrations saisonnières, la politique courante, car affectant aussi la taille des populations à risques.

Il existe plusieurs approches d'estimation des populations à risque, qui sont de nos jours pratiquées. On peut entre autre citer (WHO 2004) :

- recensement et méthodes d'énumération
- méthodes multiplicatives
- enquêtes, exemple EDS (enquête démographique et de santé) administrées avec le concours d'UNAIDS et USAIDS
- méthode par nomination
- méthode de poisson tronqué
- méthodes comportementales

2- Analyse théorique et descriptif des données

Approche théorique

Notre but est d'utiliser les différentes approches statistiques pour valider ou invalider certains des facteurs. La variable à expliquer est de type catégoriel disposant deux modalités.

Elle prend la valeur 1 si le test est positif, c'est-à-dire si l'individu est infecté, et zéro sinon.

Il existe aussi des individus au résultat final indéterminé, mais ils ne sont qu'au nombre de 4.

L'approche épidémiologique consiste à expliquer ces résultats par un ensemble de variables ou facteurs de risque. Elles sont de nature comportementale et socio démographique.

Toutes sont présentées sous forme polytomique. Une analyse multi variée sera faite en considérant les facteurs de risque comme variables latentes dans un modèle logistique. Ce dernier nous permettra de mesurer la sensibilité de la probabilité qu'un individu soit séropositif par rapport aux facteurs de risque considérés.

En effet, posons X le vecteur des facteurs de risque, β un vecteur de paramètres à estimer contenant une constante, u_i une innovation, Y_i une variable dichotomique (résultat final dans notre cas) Y_i^* une variable latente, alors le modèle logit est formulé ainsi qu'il suit :

$$Y_i = 1 \text{ si } Y_i^* \geq 0$$

$$0 \text{ sinon, avec } Y_i^* = X_i \beta + u_i, \quad i=1 \dots n \text{ le nombre d'observations.}$$

Les fonctions de densité de probabilité et de répartition du modèle logit sont respectivement :

$$f(X_i \beta) = \text{Exp}(X_i \beta) / (1 + \text{Exp}(X_i \beta))^2 ;$$

$$F(X_i \beta) = \text{Exp}(X_i \beta) / (1 + \text{Exp}(X_i \beta))$$

L'estimation du vecteur des paramètres β repose sur le maximum de vraisemblance

$$L(Y, X, \beta) = \prod (1 / (1 + \text{Exp}(X_i \beta))^{1-y_i} (\text{Exp}(X_i \beta) / (1 + \text{Exp}(X_i \beta)))^{y_i}, \quad i=1 \dots n$$

Une mesure de l'élasticité de la probabilité que $Y_i=1$ par rapport à un facteur x_{ik} est donnée

$$\text{par : } \xi = (dF(X_i \beta) / d x_{ik}) * (x_{ik} / F(X_i \beta)) = x_{ik} \beta_k / (1 + \text{Exp}(X_i \beta))$$

L'analyse sous trois angles, proposée par Liao (1994) en sera dérivée.

Continuité de l'étude

L'étude que nous menons est replicable dans d'autres zones géographiques. Elle est basée sur l'application des techniques économétriques propres aux variables catégorielles, qu'elles jouent le rôle de variable explicative ou expliquée. Il s'agit donc des cas récurrents dans les

analyses en sciences sociales. La continuité de cette étude pourra aider à éclairer les politiques globales en matière de lutte contre la pandémie du VIH/SIDA.

De plus, l'angle d'investigation par l'approche qualitative n'est pas abordé ici. Elle servirait à mieux comprendre le processus de contamination par le contrôle des facteurs qui se révèlent significatifs par le modèle ou l'approche quantitative.

Les données

Le projet EDSC III, troisième enquête démographique et de santé au Cameroun, est l'unique enquête basée sur un échantillon aléatoire et qui intègre le test du VIH/SIDA. Les deux autres enquêtes, ayant eu lieu en 1991 et 1998 ne collectaient pas cette information. Cette enquête s'effectue aussi dans d'autres pays, à travers le programme international MEASURE DHS financé par l'USAID.

Elle est coordonnée à l'échelle nationale par les instituts responsables des questions statistiques, qui bénéficient de l'expertise de ORC Macro.

Les contenus sont standardisés, mais souvent réadaptés en fonction des pays.

Trois types de questionnaires différents sont utilisés :

- Le questionnaire ménage
- Le questionnaire individuel femme
- Le questionnaire individuel mari

Du fait que les données sur le VIH soient très sensibles, le test a été fait suivant un protocole anonyme. Ainsi, il n'est pas possible de retrouver un individu ayant subi le test via les informations disponibles.

Pour les résultats du test, un fichier séparé a été constitué. Comme les caractéristiques socio démographiques sont collectées à travers les questionnaires hommes et femmes, il est possible de comparer les résultats entre pays suivant les profils.

Au Cameroun, la collecte s'est effectuée de février à août 2004.

Échantillonnage

L'échantillon de cette enquête est aréolaire et stratifié à deux degrés : l'unité primaire de sondage est la zone de dénombrement (ZD), et le deuxième degré a pour base les ménages dénombrés à l'intérieur des ZDs. En moyenne, 466 ZDs (244 urbaines contre 222 rurales) sur

17 500 ZDs ont été sélectionnées avec des probabilités proportionnelles au nombre de ménages des ZDs.

Environ 11 584 ménages (5368 urbains contre 6216 ruraux) ont été tirés au deuxième degré, et de ces ménages, un sur deux était sélectionné pour le questionnaire hommes (ages de 15-59 ans), alors que toutes les femmes de 15-49 ans étaient éligibles pour le questionnaire femmes dans l'ensemble des ménages. Environ 22 ménages étaient tirés par ZD en zone urbaine contre 28 en zone rurale.

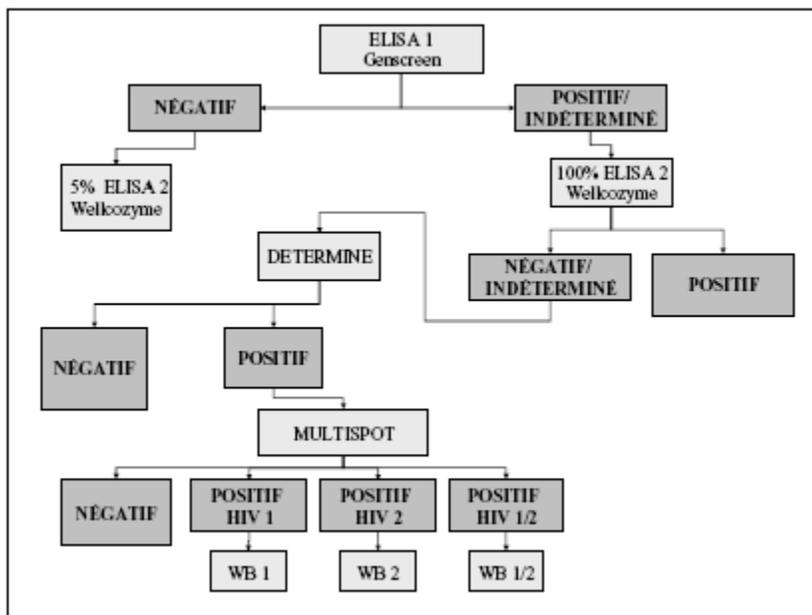
Le test du VIH a été effectué dans le sous échantillon des questionnaires hommes, et donc toutes les femmes (15-49 ans) et hommes (15-59 ans) étaient éligibles pour le test, soit au total, 5703 femmes et 5676 hommes.

Le test en Laboratoire

Les gouttes de sang collectées devaient passer deux tests : le test de dépistage via ELISA1 et le test de confirmation via ELISA 2.

En plus, 5% des échantillons négatifs ont à nouveau été testés suivant la procédure complète. Le schéma suivant en est un résumé.

Figure 2.1 : Algorithme de dépistage du VIH



Source : Centre pasteur du Cameroun, projet EDSC III

Le test des anticorps est basé sur des gouttes de sang séchées au Centre Pasteur du Cameroun, et le statut sérologique est déterminé suivant l'algorithme ci-dessus. Les gouttes de sang sont passées en ELISA 1, qui permet de séparer les individus en sérologie négative, sérologie positive et sérologie indéterminée. L'échantillon de sérologie positive ou indéterminée subit un deuxième test (ELISA 2), celui-ci permettant d'avoir 3 groupes comme précédemment : sérologie positive, négative ou indéterminée.

Les sous groupes de sérologie positive ou indéterminé subit un nouveau test (Determine, ABBOTT). L'on obtient deux groupes de sérologie positive et négative. L'échantillon positif en Determine est passe ensuite en Multispot pour une différenciation HIV-1 ou HIV-2, HIV 1 et 2 ou négatif. A ce stade les échantillons douteux ou négatifs sont rendus indéterminés, alors que ceux de réaction HIV1, HIV2 ou HIV 1 et 2 sont passés en Western Blot, d'où les résultats sont rendus positifs ou indéterminés. De plus, 5% de l'échantillon négatif au premier test subi l'algorithme complet pour un contrôle de qualité (Rapport EDSC III, 2004). Au bout de cette procédure, l'on a obtenu 4 cas indéterminés, 5.5% de cas infectés au VIH dans la tranche 15-49 ans.

3- Modèle de validation

a) Analyse descriptive

Dans cette partie, nous avons croisé les résultats du test avec des variables sélectionnées selon leur redondance dans la littérature sur le VIH et la disponibilité de données dans notre base. Le traitement statistique a été fait par le logiciel SPSS 11.5; et les résultats bruts sont en annexe. Les noms de variables retenus dans le traitement sont entre parenthèses.

Les dites variables ont été codifiées à nouveau pour répondre à nos objectifs. Par conséquent, les résultats peuvent légèrement différer de ceux du rapport de l'enquête à la base de ces données. L'analyse descriptive intègre toutes les classes d'âge. Par contre pour des raisons de comparabilité suivant le genre, nous avons exclu la tranche âgée d'au moins 50 ans des régressions, celle-ci n'étant valable que pour les hommes. On y trouve 9 infections sur 410, soit un taux de prévalence de 2.2 %.

Le niveau d'éducation (EDUCAT)

Elle est catégorielle et comporte quatre modalités : pas fréquenté- niveau primaire- niveau secondaire et niveau supérieur. Cette variable indique le niveau scolaire des personnes ayant fait le test de dépistage au VIH. Son importance se justifie par le rôle de l'information dans la vie moderne. Un individu qui ne sait ni lire, ni écrire est exclu s'il n'y a pas de moyens de communication en sa langue d'usage. On s'attend donc à ce que les plus instruits soient moins infectés que les moins scolarisés.

99.9% de l'échantillon possède cette information. On y trouve 3.6% de personnes infectées chez ceux qui n'ont pas fréquenté, 5.7% au niveau primaire, 6.0% au niveau secondaire et 4.2% du niveau supérieur. Il en ressort que les plus touchés sont ceux du secondaire, et que les moins instruits sont les moins victimes contrairement à ce que l'on croirait.

Le lieu de résidence (RESIDENC)

Cette variable indique le milieu dans lequel vit la personne. Elle est catégorielle et comporte quatre modalités, selon la cartographie spatiale existante dans la base de données : Yaoundé/Douala- Garoua/Maroua/Bafoussam/Bamenda- autres villes- milieu rural. En effet, les modes de transmissions épousent les réalités socio-économiques du milieu. L'on s'attendrait que les zones à fortes activités et qui jouissent d'une liberté de vie des membres soient plus exposées.

Le groupe des villes de Garoua, Maroua, Bafoussam et Bamenda présente le plus fort taux de prévalence (7.6%) alors que la zone rurale a le moins élevé (4.2%).

Le statut matrimonial (STATMATR)

Elle est catégorielle et compte six modalités : jamais marié- marié- concubinage- veuf (ve)- divorcé- ami(e). Le mariage est considéré comme un facteur de stabilisation. De ce fait, les non mariés seraient plus exposés que ceux qui vivent dans un foyer.

88.3% de l'échantillon a une information complète. La classe de personnes vivant avec un ami sous des toits différents a le taux le plus faible (3.0%), alors que les divorcés possèdent le plus élevé (5.9%).

La taille du ménage (TAILLEME)

Elle prend des valeurs discrètes. Plus la taille du ménage est grande, moins le responsable serait capable de mener à bien son rôle. Dans le contexte des pays en voie de développement, la taille plus élevée est aussi un indicateur de dénuement ou de dépendance. On assiste parfois à des cas où les jeunes se livrent à la prostitution pour subvenir aux besoins du ménage. Le taux le plus faible se trouve dans les familles de 6 à 9 personnes (4.9%), alors qu'il est bien élevé chez les ménages de plus de 10 personnes, soit 6.7%.

La religion (RELIGION)

Elle est catégorielle et compte sept modalités : catholique- protestant- musulman- animiste- nouvelles églises- autre église- pas de religion. Le groupe des animistes possède le plus élevé des taux (7.7%), suivi de ceux qui ne sont d'aucune religion (5.8%).

Statut à l'emploi (EMPLOI)

Cette variable permet de faire la différence entre ceux qui sont à l'emploi au moment de l'enquête, de ceux qui sont en chômage. Elle a deux modalités selon que l'individu travaille ou pas. Lorsqu'on n'a pas d'emploi, on est aussi vulnérable, et dans les conditions des PVD, il arrive que la personne se livre à des activités qui ne l'honorent pas forcément. Il n'y aurait pas une différence notable chez les deux catégories : 5.0% chez les chômeurs contre 5.6% chez ceux qui ont un emploi.

Fumeur (FUMEUR)

Cette variable permet de distinguer les fumeurs des non fumeurs. Contre toute attente, 5.4% des personnes qui ne fument pas sont séropositives, contre 3.03% chez ceux qui fument.

Sexe (SEX), âge (AGEIND) et âge au premier rapport sexuel

Le sexe prend deux valeurs selon que l'individu est une femme ou non, alors que les deux autres sont discrètes. Une représentation selon des tranches d'âge est faite sur le graphique 1 en annexe.

4% des hommes sont porteurs, contre 6.9% chez les femmes. Ceci montre que les femmes sont les plus exposées à l'infection, soit environ 2 femmes contre 1 homme. Chez les moins de 25 ans, le taux de prévalence est de 3.2%, et 8.3% chez les personnes de 25 à 34 ans contre 6.8% dans la tranche supérieur. La tranche la plus touchée est celle des jeunes en âge de transition.

Globalement, 7.3% des personnes ayant connu leur première aventure à l'âge de 20 ans sont infectées, contre 9.0% chez celles de 25 ans. En dessous de 20 ans, le taux est de 5.1% environ.

b)- Analyse multi variée

La variable dépendante étant dichotomique, le modèle logistique a été retenu.

Toutes les variables ci dessus sont des régresseurs, auxquelles s'ajoute une constante. La codification s'est faite de façon automatique via le logiciel SPSS.

Nous abordons cette partie avec deux approches : une approche d'ensemble et une approche suivant le genre. Trois modèles en sont donc dérivés.

Le tableau 3.1 donne le codage des variables nominales retenues.

Les résultats du tableau 3.2 montrent que le modèle est significatif, avec un khi carré de 210.934, pour 24 degrés de liberté. Cinq variables en sortent comme pertinentes statistiquement : le sexe, l'âge, le niveau d'éducation, le lieu de résidence, et la taille du ménage dans lequel vit l'individu.

Le test de Hosmer-Lemeshow (qui permet de tester l'hypothèse nulle selon laquelle il n'y a pas de différence entre les valeurs observées et prédites de la variable dépendante) donne une significativité de $0.439 > 0.05$. Donc l'hypothèse nulle est rejetée, ce qui montre que le modèle retenu prévoit de façon acceptable les données.

Le tableau de classification montre que même si notre modèle est fiable, les variables retenus ne permettent pas de différencier les séropositifs des séronégatifs. Ce résultat est conforme à notre objectif de départ qui consistait à identifier les variables les plus significatives pour orienter les politiques, et non d'identifier les malades.

Tableau 3.2 : résultats (estimés des coefficients) du modèle

		Variables in the Equation					95,0% C.I.for EXP(B)		
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1	RELIGION			6,606	6	,359			
	RELIGION(1)	-,041	,301	,019	1	,892	,960	,532	1,731
	RELIGION(2)	-,020	,301	,005	1	,946	,980	,544	1,767
	RELIGION(3)	-,085	,315	,073	1	,787	,918	,495	1,704
	RELIGION(4)	-,380	,454	,700	1	,403	,684	,281	1,666
	RELIGION(5)	,320	,495	,417	1	,518	1,377	,521	3,636
	RELIGION(6)	-,597	,384	2,410	1	,121	,551	,259	1,170
	STATMATR			8,236	5	,144			
	STATMATR(1)	,700	,292	5,758	1	,016	2,015	1,137	3,570
	STATMATR(2)	,721	,283	6,516	1	,011	2,057	1,182	3,578
	STATMATR(3)	,555	,297	3,475	1	,062	1,741	,972	3,119
	STATMATR(4)	,404	,415	,945	1	,331	1,497	,663	3,378
	STATMATR(5)	,740	,545	1,845	1	,174	2,096	,720	6,096
	EDUCAT			20,286	3	,000			
	EDUCAT(1)	,057	,340	,029	1	,866	1,059	,544	2,061
	EDUCAT(2)	,684	,312	4,821	1	,028	1,982	1,076	3,650
	EDUCAT(3)	,667	,307	4,717	1	,030	1,949	1,067	3,559
	RESIDENC			35,519	3	,000			
	RESIDENC(1)	,435	,135	10,407	1	,001	1,545	1,186	2,013
	RESIDENC(2)	,849	,202	17,713	1	,000	2,338	1,574	3,473
	RESIDENC(3)	,580	,113	26,301	1	,000	1,786	1,431	2,229
	AGEIND			96,813	2	,000			
	AGEIND(1)	-,964	,125	59,213	1	,000	,381	,298	,487
	AGEIND(2)	,142	,111	1,639	1	,200	1,152	,928	1,431
	TAILLEME			9,058	2	,011			
	TAILLEME(1)	-,214	,129	2,759	1	,097	,808	,628	1,039
	TAILLEME(2)	-,340	,113	9,058	1	,003	,712	,571	,888
	FUMEUR(1)	-,516	,423	1,488	1	,223	,597	,260	1,368
	EMPLOI(1)	-,175	,106	2,749	1	,097	,839	,682	1,032
	SEX(1)	,513	,098	27,221	1	,000	1,671	1,378	2,026
	Constant	-4,009	,515	60,648	1	,000	,018		

a. Variable(s) entered on step 1: RELIGION, STATMATR, EDUCAT, RESIDENC, AGEIND, TAILLEME, FUMEUR, EM SEX.

Trois pistes d'interprétation sont valables, et nous en donnons une brève description de chacune dans ce qui suit.

Effet marginal

L'effet marginal sur la probabilité d'être infecté lorsqu'on passe de la classe de base à une autre est donné par la colonne Exp(B).

Ainsi, les chances d'être infecté pour un individu n'ayant pas fait l'école est 1.06 fois celle de celui ayant atteint le niveau supérieur (classe de base), contre 1.98 fois et 1.95 fois celui du

niveau primaire ou secondaire respectivement. Une approche avec les conclusions de l'analyse descriptive permet de comprendre que les moins instruits prennent peu de risque que ceux ayant atteint le primaire ou le secondaire pour qui le risque est deux fois.

Celles des habitants des différentes catégories de villes par rapport au milieu rural sont multipliées par 1.54, 1.78 et 2.33 pour Yaoundé/Douala, autres villes et Garoua/Maroua/Bafoussam/Bamenda respectivement. Un regard particulier sur les autres villes est ainsi indispensable, plutôt qu'une concentration des ressources sur les deux grandes métropoles qui sont Yaoundé et Douala.

Le risque pour une femme d'être infectée est d'environ deux fois celui de l'homme.

Par rapport aux familles de 10 personnes ou plus, le risque diminue : il est 0.8 fois dans les moins de cinq personnes, contre 0.71 dans les 5 à 9 personnes.

Selon l'âge, il diminue dans la tranche inférieure, mais augmente dans la tranche 25-34 ans.

Prédiction des probabilités

Cet exercice consiste à estimer les probabilités d'infection suivant les profils des individus.

En effet, chaque individu (i) est considéré comme un vecteur X_i de données, où chaque élément est une valeur prise par une variable du modèle. Ces éléments sont remplacés dans l'expression $P(Y=1) = \text{Exp}(X_i \beta) / (1 + \text{Exp}(X_i \beta))$.

Effet marginal sur la probabilité d'un évènement

Pour les variables continues, Il est capturé par

$dP(Y=1)/dx_k = \beta_k * \text{Exp}(X_i \beta) / (1 + \text{Exp}(X_i \beta))^2$ ou d est l'opérateur de dérivée partielle.

Par contre pour les variables catégorielles, il faut reconsidérer les différences de probabilités à partir de l'analyse descriptive.

Toutefois, on obtient des résultats semblables avec la dérivée partielle ci-dessus pour les différents types de variables (Liao, 1994).

Le fait que la variable sexe soit significative dans le modèle d'ensemble nous permet de penser que les femmes et les hommes n'ont pas le même degré d'exposition à l'infection.

Ainsi, les comportements pourraient différer suivant le sexe, impliquant une approche genre dans les politiques à mettre en œuvre. Ce constat nous conduit à envisager un modèle de validation séparé, d'un côté pour les hommes, et de l'autre pour les femmes. Les résultats sont contenus dans les tableaux 2.3 et 3.3 respectivement en annexes 2 et 3.

Le modèle répliqué suivant le genre donne des renseignements sur la sensibilité par rapport au sexe. Il en ressort que les variables de lieu de résidence, l'âge et le niveau d'éducation sont significatives chez les femmes, et à 10%, la religion, le statut matrimonial et la taille du ménage le sont aussi. La statistique de Hosmer-Lemeshow est de $0.145 > 0.05$; ce qui signifie que le modèle est acceptable, mais ne peut permettre une différenciation de la sérologie.

Chez les hommes, les variables liées à l'âge, la taille du ménage et le lieu de résidence sont des variables significatives. Par contre, le niveau d'éducation ne l'est pas. Une question peut ainsi être posée, à savoir ce qui soutiendrait l'importance de la taille du ménage dans les attitudes à haut risque chez les hommes?

La statistique de Hosmer-Lemeshow est de $0.076 > 0.05$; donc le modèle est acceptable, mais ne peut permettre une différenciation de la sérologie.

c)- Implications sur les politiques publiques

Notre modèle nous permet de focaliser notre attention sur les cinq facteurs les plus significatifs. L'avantage par rapport à une analyse descriptive est ainsi révélé. En effet si la dernière permet de prendre des décisions globales (programmes), le modèle permet de choisir des actions plus spécifiques et immédiates via des projets bien ciblés.

Le couplage de cette approche, qui est purement quantitative avec une recherche qualitative permettrait de mieux cerner les contours cachés, et donc de mieux circonscrire des actions de court et de long terme dans la lutte contre la montée de l'infection.

Des politiques en paliers sont donc nécessaires dans la lutte contre la pandémie du VIH au Cameroun. Le phénomène touche différemment les hommes et les femmes selon l'âge et le milieu dans lequel la personne vit.

Conclusion

La question à laquelle nous voulions répondre, à savoir « quels sont les facteurs de risques valides » nous a permis de passer en revue la littérature existante sur la question. La littérature est très riche en matière de modèles mathématiques des différentes phases de la maladie, mais des études de cas sont moins abondantes.

Les données que nous utilisons sont issues de l'enquête démographique et de santé au Cameroun, réalisée en 2004. La population concernée est constituée d'hommes et femmes sexuellement actifs: 15-49 ans pour les femmes, et 15-59 ans pour les hommes.

L'approche épidémiologique de la question consiste à identifier un panel de facteurs pouvant expliquer la propension à l'infection. Aussi, le sexe, l'âge, le statut matrimonial, le niveau d'éducation, le zone de résidence, la situation par rapport à l'emploi, le nombre partenaires sexuels, la prise de drogue etc., constituent des facteurs candidats.

Notre approche a consisté en l'étude de l'ensemble de l'échantillon, puis à en faire deux modèles séparés pour les hommes et les femmes, soit trois modèles obtenus.

Il en ressort que le modèle commun est globalement significatif. Le sexe, l'âge, le lieu de résidence, le niveau d'éducation et la taille du ménage sont significatives.

Les variables lieu de résidence, l'âge et le niveau d'éducation sont significatives chez les femmes.

À 10%, la religion, le statut matrimonial et la taille du ménage le sont aussi.

Chez les hommes, l'âge, la taille du ménage et le lieu de résidence sont significatives. Par contre, le niveau d'éducation ne l'est pas.

Dans les trois cas, les modèles sont acceptables, mais ne peuvent être utilisés pour différencier les individus suivant leur sérologie, étant données les observations disponibles. Ceci reste conforme avec notre objectif de départ, qui visait à permettre d'extraire les variables d'action pour les politiques gouvernementales. Un rapprochement avec les résultats de l'analyse descriptive montre une complémentarité, et donc suggère que des actions soient prises suivant une approche à la fois globale et ciblée selon le genre.

Une recherche qualitative est aussi nécessaire pour une meilleure compréhension des aspects non saisissables par l'approche quantitative, à la fois pour corroborer avec les résultats de cette dernière, soit pour l'orienter, ou alors pour aider à l'élaborer (Rossman and Bruce, 1985).

Bibliographie

- 1- Alban T., *Econométrie des variables qualitatives*, France, Dunod, Paris, 2000.
- 2- Anderson, R.M , “The role of mathematical models in the study of HIV transmission and the epidemiology of AIDS”, *Journal of Acquired Immune Deficiency Syndromes*, vol.1, 1988a, 241-256.
- 3- Anderson, R.M , “The epidemiology of HIV infection : variable incubation plus infectious periods and heterogeneity in sexual activity”, *Journal of the Royal Statistical Society*, vol. 151, 1988b, 66-93.
- 4- Anderson, R.M et May, “Epidemiological parameters of HIV transmission”, *Nature*, vol.333, 1988, 514-519
- 5- Anderson, R.M , “Mathematical and statistical studies studies of the epidemiology of HIV AIDS”, *Journal of the Royal Statistical Society*, vol.3, 1989, 333-346
- 6- Cox, D.R and Oakes, D., *Analysis of survival data*, Chapman and Hall, London, 1984
- 7- Cox, D.R and Oakes, D., “Regression models and life-tables”, *Journal of the Royal Statistical Society*, vol. 34, 1972, 1987-220.
- 8- Brookmeyer R., and Liao, J., “Statistical modeling of the AIDS epidemic for forecasting health care needs”, *Biometrics*, vol.46, 1990a, 1151-1163.
- 9- Brookmeyer, R. & Mitchel, H.G., *AIDS epidemiology : a quantitative approach*, Oxford University press, 1994.
- 10- Breslow at al., *Statistical methods in cancer research*, vol.1, the analysis of case-control studies, France, Lyon, International Agency for Research on Cancer, 1990.
- 11- Darby, S.C et al., “Time from infection with HIV to onset of AIDS in patients with hemophilia in the U.K”, *Statistics in medicine*, vol.9, 1990, 681-689.
- 12- Darrow et al., “Risk factors for human immunodeficiency virus infections in homosexual men”, *American Journal of Public Health*, vol.77, 479-483.
- 13- De Grutola, V., Lagakas, S.W., “Modeling the progression of HIV infection”, *Journal of the American Statistical Association*, vol.86, 1991, 569-577.
- 14- De Grutola, et al.. “Assessing and modeling heterosexual spread of the human immunodeficiency virus in the United States ”, *Review of Infectious Diseases*, vol.10, 1988, 138-150.
- 15- De Vincenzi, I., “Heterosexual transmission of HIV (letter) ”, *Journal of the American Medical Association*, vol.267, 1992, 1919
- 16- Gretchen B. Rossman, Bruce L. Wilson, “ Numbers and words: combining quantitative and qualitative methods in a single large-scale evaluation study”, *Evaluation Review*, vol. 9 N0 5, October 1985, 627-643.
- 17- Institut National de la Statistique, *Enquête démographique et de santé Cameroun*, Yaoundé, 2005
- 18- UNAIDS, *Estimating the size of populations at risk for HIV*, 2003.

Annexes

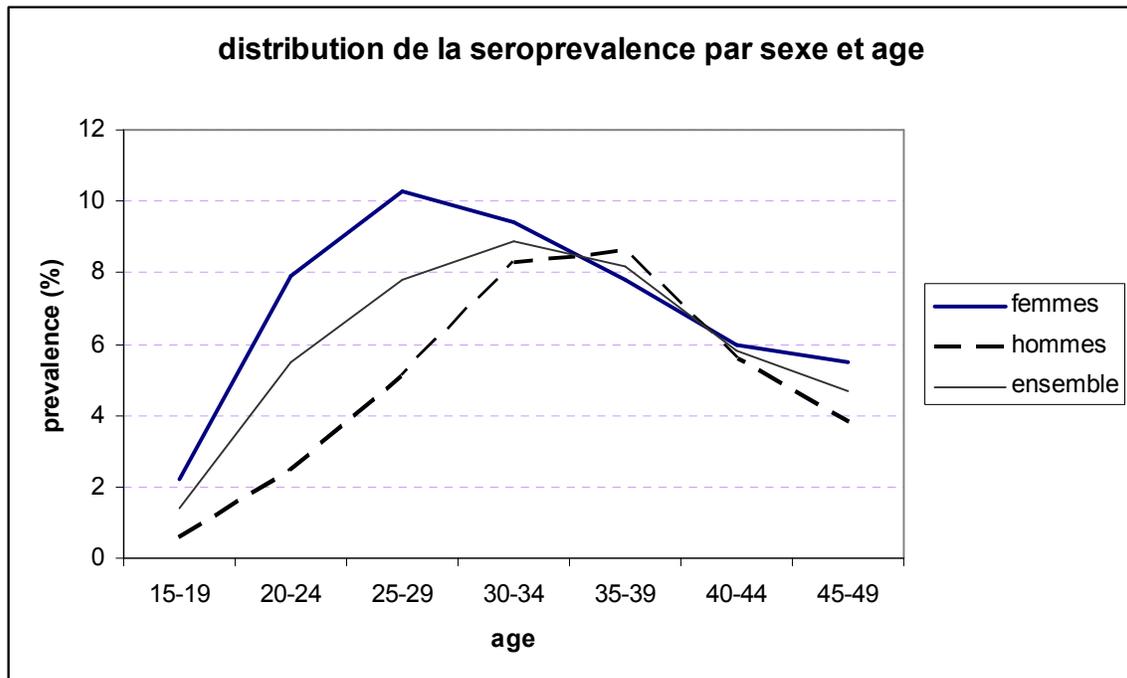
Annexe1 : Prévalence par tranche d'âge et par sexe

Tableau 1.1 : Séroprévalence

Age	Femmes		Hommes		Ensemble	
	positifs(%)	effectif	positifs(%)	effectif	positifs(%)	effectif
15-19	2,2	1297	0,6	1209	1,4	2507
20-24	7,9	1104	2,5	917	5,5	2021
25-29	10,3	868	5,1	786	7,8	1654
30-34	9,4	686	8,3	602	8,9	1288
35-39	7,8	519	8,6	447	8,2	966
40-44	6	416	5,6	369	5,8	785
45-49	5,5	337	3,8	342	4,7	679
50-54	-	-	2,5	255	-	-
55-59	-	-	1	197	-	-

Source : Institut National de la Statistique du Cameroun, 2004

Graphique 1.1 : séroprévalence en tranche d'âge



Annexe 2 : Résultats d'ensemble

Tableau 2.1: Résumé des observations prises en compte dans le modèle

Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	8948	74,2
	Missing Cases	3117	25,8
	Total	12065	100,0
Unselected Cases		0	,0
Total		12065	100,0

a. If weight is in effect, see classification table for the total number of cases.

Il y a perte de 25,8% de données du fait d'informations manquantes sur les individus concernés.

Tableau 2.2 : classification en catégories

Classification Table^a

Observed			Predicted		Percentage Correct
			resultat du test		
			negatif	seropositif	
Step 1	resultat du test	negatif	8450	0	100,0
		seropositif	498	0	,0
Overall Percentage					94,4

a. The cut value is ,500

Le modèle permet de prévoir parfaitement les cas séronégatifs, mais aucun cas séropositif.

Tableau 2.3 : Codification automatique des variables (regresseurs)

Categorical Variables Codings

		Frequency	Parameter coding					
			(1)	(2)	(3)	(4)	(5)	(6)
Religion	Catholique	3419	1,000	,000	,000	,000	,000	,000
	Protestant	3089	,000	1,000	,000	,000	,000	,000
	Musulman	1496	,000	,000	1,000	,000	,000	,000
	Animiste	195	,000	,000	,000	1,000	,000	,000
	autre religion	85	,000	,000	,000	,000	1,000	,000
	pas de religion	450	,000	,000	,000	,000	,000	1,000
	Eglises Reveillées	214	,000	,000	,000	,000	,000	,000
statut matrimonial	jamais marié	2101	1,000	,000	,000	,000	,000	
	Marrié	4582	,000	1,000	,000	,000	,000	
	concubinage	1470	,000	,000	1,000	,000	,000	
	veuf(ve)	244	,000	,000	,000	1,000	,000	
	Divorcé	80	,000	,000	,000	,000	1,000	
niveau education	copain	471	,000	,000	,000	,000	,000	
	pas fréquenté	1386	1,000	,000	,000			
	primaire	3501	,000	1,000	,000			
	secondaire	3746	,000	,000	1,000			
lieu residence	superieur	315	,000	,000	,000			
	Yaounde/Douala	1637	1,000	,000	,000			
	Garoua/Maroua/Ba foussa/Bamenda	404	,000	1,000	,000			
	Autres villes	2319	,000	,000	1,000			
age individuel	Rural	4588	,000	,000	,000			
	moins de 25 ans	4041	1,000	,000				
	25-34 ans	2645	,000	1,000				
taille du menage	35-49 ans	2262	,000	,000				
	1-4 personnes	2299	1,000	,000				
	5-9 pers.	4486	,000	1,000				
statut de l'emploi	au moins 10 pers.	2163	,000	,000				
	chomeur	3608	1,000					
fumeur	occupé	5340	,000					
	oui	191	1,000					
sexe de l'individu	non	8757	,000					
	femme	4743	1,000					
	homme	4205	,000					

La dernière catégorie de chaque variable est la base dans ce schéma de codification. Ainsi, les analyses pour chacune des variables sont faite par comparaison à cette base.

Les catégories de base sont donc respectivement: églises de réveille, copain, supérieur, au moins 10 personnes, 35-49 ans, occupé, non fumeur, et homme.

Annexe 3: Resultats pour hommes

Tableau 3.1: Résumé des observations prises en compte dans le modèle

Case Processing Summary

Unweighted Cases ^b	N	Percent
Selected Cases ^a Included in Analysis	4205	34,9
Missing Cases	1833	15,2
Total	6038	50,0
Unselected Cases	6027	50,0
Total	12065	100,0

a. The category variable sexe de l'individu is constant for all selected cases. Since a constant was requested in the model, it will be removed from the analysis.

b. If weight is in effect, see classification table for the total number of cases.

Tableau 3.2 : classification en catégories

Classification Table

Observed	Predicted					
	Selected Cases ^a			Unselected Cases ^{b,c}		
	resultat du test		Percentage	resultat du test		Percentage
	negatif	seropositif	Correct	negatif	seropositif	Correct
Step 0 resultat du test negatif	4024	0	100,0	4426	0	100,0
seropositif	181	0	,0	317	0	,0
Overall Percentage			95,7			93,3

a. Selected cases Sex of household member EQ 1

b. Unselected cases Sex of household member NE 1

c. Some of the unselected cases are not classified due to either missing values in the independent variables or with values out of the range of the selected cases.

d. Constant is included in the model.

e. The cut value is ,500

Comme dans le modèle d'ensemble, il n'est pas possible d'avoir une prédiction des cas séropositifs.

Tableau 3.3 : estimés des coefficients des regressseurs

		Variables in the Equation						95,0% C.I.for EXP(B)	
		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Step 1 ^a	RELIGION			3,031	6	,805			
	RELIGION(1)	-,238	,451	,279	1	,597	,788	,326	1,907
	RELIGION(2)	-,338	,453	,556	1	,456	,713	,294	1,734
	RELIGION(3)	-,097	,474	,042	1	,838	,907	,358	2,297
	RELIGION(4)	,273	,617	,196	1	,658	1,314	,392	4,406
	RELIGION(5)	-,082	,863	,009	1	,924	,921	,170	4,996
	RELIGION(6)	-,454	,567	,641	1	,423	,635	,209	1,929
	STATMATR			3,001	5	,700			
	STATMATR(1)	,485	,449	1,163	1	,281	1,624	,673	3,919
	STATMATR(2)	,336	,437	,592	1	,442	1,399	,595	3,292
	STATMATR(3)	,515	,457	1,267	1	,260	1,673	,683	4,101
	STATMATR(4)	-,192	,723	,071	1	,790	,825	,200	3,407
	STATMATR(5)	,749	,859	,760	1	,383	2,115	,393	11,396
	EDUCAT			3,288	3	,349			
	EDUCAT(1)	,160	,439	,132	1	,716	1,173	,496	2,776
	EDUCAT(2)	,520	,360	2,092	1	,148	1,682	,831	3,403
	EDUCAT(3)	,472	,349	1,830	1	,176	1,604	,809	3,179
	RESIDENC			17,792	3	,000			
	RESIDENC(1)	,423	,224	3,553	1	,059	1,526	,983	2,370
	RESIDENC(2)	1,083	,317	11,670	1	,001	2,952	1,586	5,495
	RESIDENC(3)	,616	,186	10,903	1	,001	1,851	1,284	2,668
	AGEIND			61,905	2	,000			
	AGEIND(1)	-1,722	,232	55,280	1	,000	,179	,113	,281
	AGEIND(2)	-,085	,169	,251	1	,617	,919	,659	1,281
	TAILLEME			7,056	2	,029			
	TAILLEME(1)	-,490	,219	4,991	1	,025	,612	,398	,942
	TAILLEME(2)	-,415	,178	5,425	1	,020	,660	,466	,936
	FUMEUR(1)	-,715	,728	,964	1	,326	,489	,117	2,039
	EMPLOI(1)	-,179	,174	1,055	1	,304	,836	,595	1,176
	Constant	-3,088	,698	19,551	1	,000	,046		

a. Variable(s) entered on step 1: RELIGION, STATMATR, EDUCAT, RESIDENC, AGEIND, TAILLEME, FUMEUR, EMPLOI.

De ce tableau sont extraites les variables significatives. La colonne sig. Est utilisée, en se fixant un seuil 5% ou 10% par exemple. La variable est significative si la valeur de cette colonne est inférieure au seuil fixe.

Annexe 4: Resultats pour femmes

Tableau 4.1: Résumé des observations prises en compte dans le modèle

Case Processing Summary

Unweighted Cases ^b		N	Percent
Selected Cases ^a	Included in Analysis	4743	39,3
	Missing Cases	1284	10,6
	Total	6027	50,0
Unselected Cases		6038	50,0
Total		12065	100,0

a. The category variable sexe de l'individu is constant for all selected cases. Since a constant was requested in the model, it will be removed from the analysis.

b. If weight is in effect, see classification table for the total number of cases.

Tableau 4.2 : Classification en catégories

Classification Table

		Predicted					
		Selected Casés ^a			Unselected Casés ^{b,c}		
		resultat du test		Percentage	resultat du test		Percentage
		negatif	seropositif	Correct	negatif	seropositif	Correct
Observed							
Step 1 resultat du test	negatif	4426	0	100,0	4024	0	100,0
	seropositif	317	0	,0	181	0	,0
Overall Percentage				93,3			95,7

a. Selected cases Sex of household member EQ 0

b. Unselected cases Sex of household member NE 0

c. Some of the unselected cases are not classified due to either missing values in the independent variable with values out of the range of the selected cases.

d. The cut value is ,500

Comme dans le modèle d'ensemble, il n'est pas possible d'avoir une prédiction des cas séropositifs.

Tableau 4.3 : estimés des coefficients des regresseurs

		Variables in the Equation					95,0% C.I. for EXP(B)		
Step		B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
1 ^a	RELIGION			11,303	6	,079			
	RELIGION(1)	,084	,407	,042	1	,837	1,087	,489	2,416
	RELIGION(2)	,164	,407	,162	1	,687	1,178	,531	2,613
	RELIGION(3)	-,089	,426	,043	1	,835	,915	,397	2,110
	RELIGION(4)	-1,019	,712	2,049	1	,152	,361	,089	1,457
	RELIGION(5)	,566	,624	,824	1	,364	1,762	,519	5,983
	RELIGION(6)	-,702	,525	1,786	1	,181	,496	,177	1,387
	STATMATR			9,404	5	,094			
	STATMATR(1)	,797	,385	4,287	1	,038	2,219	1,043	4,718
	STATMATR(2)	,934	,372	6,294	1	,012	2,544	1,227	5,276
	STATMATR(3)	,577	,393	2,156	1	,142	1,780	,824	3,843
	STATMATR(4)	,681	,519	1,725	1	,189	1,976	,715	5,459
	STATMATR(5)	,789	,709	1,240	1	,266	2,201	,549	8,828
	EDUCAT			18,406	3	,000			
	EDUCAT(1)	,796	,745	1,142	1	,285	2,216	,515	9,540
	EDUCAT(2)	1,503	,728	4,260	1	,039	4,497	1,079	18,746
	EDUCAT(3)	1,477	,726	4,141	1	,042	4,379	1,056	18,158
	RESIDENC			19,310	3	,000			
	RESIDENC(1)	,482	,170	8,046	1	,005	1,619	1,161	2,259
	RESIDENC(2)	,678	,264	6,586	1	,010	1,969	1,174	3,305
	RESIDENC(3)	,568	,143	15,750	1	,000	1,764	1,333	2,335
	AGEIND			41,887	2	,000			
	AGEIND(1)	-,560	,158	12,593	1	,000	,571	,419	,778
	AGEIND(2)	,340	,148	5,260	1	,022	1,405	1,051	1,878
	TAILLEME			4,695	2	,096			
	TAILLEME(1)	-,043	,162	,071	1	,789	,958	,697	1,315
	TAILLEME(2)	-,279	,147	3,598	1	,058	,757	,567	1,009
	FUMEUR(1)	-,399	,524	,579	1	,447	,671	,241	1,874
	EMPLOI(1)	-,175	,133	1,715	1	,190	,840	,647	1,091
	Constant	-4,823	,923	27,328	1	,000	,008		

a. Variable(s) entered on step 1: RELIGION, STATMATR, EDUCAT, RESIDENC, AGEIND, TAILLEME, FUMEUR, EMPLOI.

De ce tableau sont extraites les variables significatives. La colonne sig. Est utilisée, en se fixant un seuil 5% ou 10% par exemple. La variable est significative si la valeur de cette colonne est inférieure au seuil fixe.