

Estimation de la distribution des temps d'infection par le VIH à partir des données longitudinales de marqueurs virologiques de séroconversion

Cécile Sommen, Ahmadou Alioum, Daniel Commenges

► **To cite this version:**

Cécile Sommen, Ahmadou Alioum, Daniel Commenges. Estimation de la distribution des temps d'infection par le VIH à partir des données longitudinales de marqueurs virologiques de séroconversion. 41èmes Journées de Statistique, SFdS, Bordeaux, 2009, Bordeaux, France, France. 2009. <inria-00386571>

HAL Id: inria-00386571

<https://hal.inria.fr/inria-00386571>

Submitted on 22 May 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ESTIMATION DE LA DISTRIBUTION DES TEMPS D'INFECTION PAR LE VIH À PARTIR DES DONNÉES LONGITUDINALES DE MARQUEURS VIROLOGIQUES DE SÉROCONVERSION

Cécile Sommen¹, Ahmadou Alioum¹, Daniel Commenges¹

¹ *INSERM, U897, Université Victor Segalen Bordeaux 2, 146 rue Léo Saignat, 33076 Bordeaux Cedex, France.*

Résumé: Depuis les années 1990, de nombreux travaux portent sur l'étude de l'évolution des anticorps anti-VIH menant à des tests permettant de distinguer les infections récentes des infections déjà bien établies à partir d'un seul échantillon de sérum. L'incidence peut alors être estimée à partir de la relation entre la prévalence, l'incidence et la durée de l'infection récente ("période fenêtre"). Cependant, de récents travaux ont montré les limites de cette approche dues essentiellement à une grande variabilité de la "période fenêtre". Nous proposons une approche alternative qui consiste à estimer la distribution du temps d'infection basée sur la valeur des marqueurs virologiques au moment où l'infection est découverte pour la première fois. Dans un premier temps, un modèle pour l'évolution des marqueurs est spécifié et estimé à partir de mesures répétées de marqueurs virologiques de séroconversion. Les paramètres du modèle sont estimés à partir des données d'une cohorte de patients inclus pendant la primo-infection. Dans un second temps, nous utilisons ce modèle pour estimer la distribution des temps d'infection pour les sujets nouvellement diagnostiqués VIH+ et reportés dans le système de surveillance des diagnostics VIH en France.

Mots clés : Incidence à VIH ; infections récentes ; système de surveillance ; données longitudinales ; modèle mixte

Abstract: In the last decade, interest has been focused on HIV antibody assays and testing strategies that could distinguish recent infections from established infection in a single serum sample. Incidence estimates are obtained by using the relationship between prevalence, incidence and duration of recent infection ("window period"). However, recent works demonstrated limitations of this approach due to variability of the "window period". We propose an alternative approach which consists in estimating the distribution on infection times based on serological marker values at the moment when the infection is first discovered. We propose a model based on the repeated measurements of virological markers of seroconversion for the marker trajectory. The parameters of the model are estimated using data from a cohort of HIV infected patients enrolled during primary infection. We use this model to estimate the distribution of infection times for newly HIV diagnosed subjects reported in the French HIV surveillance system.

Keywords: HIV incidence; recent infections; surveillance system; longitudinal data; mixed model

1 Introduction

Comprendre la dynamique de l'infection à VIH est essentiel pour contrôler l'épidémie du SIDA et prévoir les besoins de prise en charge. Cependant, pour la majorité des individus infectés par le VIH, la date d'infection est inconnue. L'utilisation de marqueurs sérologiques permettant d'estimer le temps écoulé depuis la séroconversion a été envisagée. En 1994, Dubin *et al.* ont proposé une méthode permettant d'estimer le temps depuis l'infection pour des individus infectés par le VIH en utilisant les valeurs des lymphocytes T CD4+. Toutefois, l'utilisation de ce marqueur n'était pas satisfaisante pour l'estimation de la date d'infection. Dans la dernière décennie, des travaux utilisant des marqueurs virologiques de séroconversion ont été développés (Brookmeyer *et al.*, 1995 ; Janssen *et al.*, 1998). Cependant, ces méthodes permettent seulement de distinguer les infections récentes des infections déjà bien établies afin d'estimer l'incidence à VIH, définie comme le nombre de nouvelles infections pendant une période de temps donnée. Brookmeyer et Quinn (1995) ont proposé une méthode consistant à tester tous les échantillons négatifs aux anticorps anti-VIH d'une enquête transversale pour détecter l'antigène p24, révélateur d'une infection dans les 14 jours. Cependant, la courte fenêtre de pré-séroconversion et la faible incidence rendaient l'estimation de cette période incertaine, et celle de l'incidence peu fiable. En 1998, Janssen *et al.* ont proposé la première approche basée sur un test à sensibilité atténuée, appelée STARHS (algorithme sérologique pour déterminer une séroconversion récente). Un résultat positif avec un test sensible, et négatif avec un test désensibilisé permettait de conclure à une infection récente. D'autres tests ont été développés pour améliorer la performance de cette approche. Cependant, des limites de cette approche ont été montrées, soulignant en particulier la variabilité de la "période fenêtre" qui définit la longueur de la période dite "récente". Le but de notre travail est de développer une méthode utilisant les données virologiques, aujourd'hui utilisées pour détecter les infections récentes, afin d'estimer, à partir d'un échantillon de sérum, la date d'infection pour un individu infecté. L'estimation de ces dates d'infection pour les individus infectés et reportés dans le système de surveillance des diagnostics de séropositivité permettra d'estimer l'incidence à VIH.

2 Modèle pour l'évolution du marqueur

Notre approche consiste dans un premier temps à déterminer un modèle permettant de décrire l'évolution des marqueurs virologiques de séroconversion à partir de la date d'infection. La dynamique de ces marqueurs a été peu étudiée. Les paramètres du modèle sont estimés à partir de mesures répétées de marqueurs virologiques de séroconversion. Plusieurs marqueurs peuvent être considérés mais dans ce travail, nous nous limiterons à un marqueur.

2.1 Spécification du modèle

Soit $Y_i = (Y_{i1}, \dots, Y_{in_i})^T$ le vecteur des valeurs du marqueur où Y_{ij} est la valeur du marqueur au temps t_{ij} pour l'observation j ($j = 1, \dots, n_i$) de l'individu i , $i = 1, \dots, n$. La date d'infection u_i est inconnue ; la seule information disponible est la dernière date de sérologie négative (l_i) et la première date de sérologie positive (r_i). Nous supposons que u_i est distribuée selon une uniforme dans $[l_i, r_i]$. Nous considérons un modèle mixte à effet aléatoire pour les Y_{ij} défini par :

$$\begin{cases} Y_{ij} = g_{\mathbf{b}}(t_{ij} - u_i, \mathbf{a}_i) + \epsilon_{ij}, i = 1, \dots, n, j = 1, \dots, n_i \\ \epsilon_i \sim N(\mathbf{0}, \sigma_e^2 \mathbf{1}) \\ \mathbf{a}_i \sim N(\mathbf{0}, D) \\ u_i \sim U[l_i, r_i] \end{cases}$$

Plusieurs choix sont possibles pour la fonction $g_{\mathbf{b}}$. Nous considérons dans un premier temps les deux fonctions suivantes :

- $g_{\mathbf{b}}(t, \mathbf{a}_i) = \frac{b_1 + a_{i1}}{1 + (b_2 + a_{i2})e^{(-t/(b_3 + a_{i3}))}}$
- $g_{\mathbf{b}}(t, \mathbf{a}_i) = (b_1 + a_{i1})e^{(b_2 + a_{i2})(1 - e^{-(b_3 + a_{i3})t})}$

avec $\mathbf{b} = (b_1, b_2, b_3)^T$ et $\mathbf{a}_i = (a_{i1}, a_{i2}, a_{i3})^T$.

Les paramètres à estimer sont les paramètres \mathbf{b} et D associés à la fonction $g_{\mathbf{b}}$ et la variance σ_e^2 .

2.2 Vraisemblance

La fonction de densité conditionnelle de Y_i sachant u_i et \mathbf{a}_i s'écrit :

$$f(Y_i | \mathbf{a}_i, u_i) = \frac{1}{(2\pi\sigma_e^2)^{n_i/2}} \cdot \exp\left(-\frac{1}{2\sigma_e^2} \cdot G_i^T G_i\right)$$

avec

$$G_i = \begin{pmatrix} Y_{i1} - g_{\mathbf{b}}(t_{i1} - u_i, \mathbf{a}_i) \\ \vdots \\ Y_{in_i} - g_{\mathbf{b}}(t_{in_i} - u_i, \mathbf{a}_i) \end{pmatrix}$$

La contribution à la vraisemblance de l'individu i peut alors s'écrire :

$$L_i = \int_{-\infty}^{\infty} \int_{l_i}^{r_i} f(Y_i | \mathbf{a}_i, u_i) f(u_i) f(\mathbf{a}_i) du_i d\mathbf{a}_i$$

On en déduit l'expression de la vraisemblance totale :

$$L = \prod_{i=1}^n L_i$$

L'estimation des paramètres du modèle se fait à partir des données $(l_i, r_i, Y_{i1}, \dots, Y_{in_i})$, $i = 1, \dots, n$ d'une cohorte de sujets infectés par le VIH et inclus dans la phase de la primo-infection.

2.3 Estimation des effets aléatoires individuels

Pour évaluer la qualité d'ajustement du modèle proposé aux données individuelles, l'estimation des effets aléatoires est nécessaire. Une approche de Bayes empirique peut être utilisée pour estimer les effets aléatoires :

$$\hat{\mathbf{a}}_i = E[\mathbf{a}_i | Y_i] = \int_{-\infty}^{\infty} \mathbf{a}_i f(\mathbf{a}_i | Y_i) d\mathbf{a}_i$$

où

$$f(\mathbf{a}_i | Y_i) = L_i^{-1} \int_{l_i}^{r_i} f(Y_i | \mathbf{a}_i, u_i) f(u_i) f(\mathbf{a}_i) d\mathbf{a}_i du_i$$

Cependant, il est difficile de calculer cette intégrale. Nous estimons alors les effets aléatoires par le mode *a posteriori* :

$$\hat{\mathbf{a}}_i = \underset{\mathbf{a}_i}{\operatorname{argmax}} \int_{l_i}^{r_i} f(Y_i | \mathbf{a}_i, u_i) \times f(\mathbf{a}_i) \times f(u_i) du_i$$

3 Estimation de la distribution *a posteriori* de la date d'infection

Dans cette partie, on suppose que l'on observe une seule valeur du marqueur Y_i et que la distribution de Y_i est connue et définie par :

$$\begin{cases} Y_i = g_{\mathbf{b}}(t_i - u_i, \mathbf{a}_i) + \epsilon_i \\ \epsilon_i \sim N(0, \hat{\sigma}_e^2) \\ \mathbf{a}_i \sim N(\mathbf{0}, \hat{D}) \end{cases}$$

avec $\hat{\mathbf{b}}$, $\hat{\sigma}_e^2$ et \hat{D} estimés dans la première étape. La deuxième étape consiste à estimer la distribution *a posteriori* des temps d'infection u_i connaissant la valeur du marqueur Y_i à la date de première sérologie positive.

En appliquant la formule de Bayes, la fonction de densité de la date d'infection *a posteriori* s'écrit :

$$f(u_i|Y_i) = L_i^{-1} \int_{-\infty}^{\infty} \hat{f}(Y_i|u_i, \mathbf{a}_i) f(u_i) \hat{f}(\mathbf{a}_i) d\mathbf{a}_i$$

Nous appliquons cette méthode sur les sujets nouvellement diagnostiqués VIH et reportés dans le système de surveillance des diagnostics VIH en France. Les marqueurs disponibles sont ceux utilisés par Barin *et al.* (2005) dont le test est utilisé pour détecter les infections récentes en France.

4 Conclusion

L'estimation de la date d'infection des sujets reportés dans le système de surveillance sur une période donnée nous permettra de déterminer combien de sujets ont été infectés dans cette période. Contrairement aux méthodes faisant intervenir une "période fenêtre" moyenne, le fait d'estimer la distribution de la date d'infection permet de contourner le problème de variabilité de cette "période fenêtre". Toutefois, notre approche repose sur un choix judicieux du modèle d'évolution des marqueurs.

Bibliographie

- [1] Dubin, N. ; Berman, S. ; Marmor M. *et al.* (1994) Estimation of time since infection using longitudinal disease-marker data, *Statistics in Medicine*, 15, 231–44.
- [2] Brookmeyer, R. et Quinn, T.C. (1995) Estimation of current HIV incidence rates from a cross-sectional survey using early diagnostic tests, *Am J Epidemiol*, 141, 166–72.
- [3] Janssen, R.S. ; Satten G.A. ; Stramer S.L. *et al.* (1998) New Testing Strategy to Detect Early HIV-1 Infection for Use in Incidence Estimates and for Clinical and Prevention Purposes, *JAMA*, 280, 42–48.
- [4] Barin, F ; Meyer, L. ; Lancar, R. *et al.* (2005) Development and Validation of an Immunoassay for Identification of Recent Human Immunodeficiency Virus Type 1 Infections and Its Use on Dried Serum Spots, *Journal of Clinical Microbiology*, 43, 4441–4447.