Sur la reproductivité nette R_0 dans les modèles d'activité sexuelle pour les épidémies de VIH/SIDA: exemple du Yunnan en Chine

Nicolas BACAËR* Xamxinur ABDURAHMAN† YE Jianli‡
Pierre AUGER§

Math. Biosci. Eng. 4 (2007) 595-607
hal : 01366423
Traductions : [ar, de, es, it, ja, nl, pt, ru, zh], [html]

Résumé

L'hétérogénéité dans les comportements sexuels joue un rôle important dans la propagation du VIH. Un modèle mathématique à base d'équations différentielles ordinaires a été proposé en 1986 pour prendre en compte la distribution de l'activité sexuelle. En supposant un mélange proportionnel, la reproductivité nette R_0 qui fixe le seuil épidémique est proportionnelle à $M + V/M$, où M est la moyenne et V la variance de la distribution. On remarque ici que cette distribution théorique est différente de celle que l'on obtient dans les enquêtes de comportements au sujet du nombre de partenaires sexuels pendant une durée τ. Cette dernière est une loi mixte de Poisson dont la moyenne m et la variance v sont telles que $M = m/\tau$ et $V = (v - m)/\tau^2$. Donc $M + V/M = (m + v/m - 1)/\tau$. De cette manière, on renforce le lien entre la théorie et les données dans les modèles d'activité sexuelle pour les épidémies de VIH/SIDA. Comme exemple, on considère les données concernant des travailleuses du sexe et leurs clients dans le Yunnan, en Chine. On trouve une borne supérieure pour la moyenne géometrique des probabilités de transmission dans ce contexte.

*Institut de Recherche pour le Développement, 32 avenue Henri Varagnat, 93143 Bondy cedex, France, nicolas.bacaer@ird.fr

†Collège de mathématiques et sciences des systèmes, Université du Xinjiang, 14 Shengli Lu, Urumqi, 830046, Chine, axamxi@xju.edu.cn

‡Centre national pour la santé maternelle et infantile, Département de la gestion de l'information, 13 Dong Tu Cheng Lu, District de Chang Yang, Pékin, 100013, Chine, yejianli@chinawch.org.cn

§pierre.auger@ird.fr

1 Introduction

En 2002, des études comportementales ont été menées sur huit cents hommes âgés de 18 à 50 ans et huit cents travailleuses du sexe dans le Yunnan et le Sichuan, deux provinces du sud-est de la Chine. Ces études faisaient partie d'un projet sino-britannique de prévention contre le VIH. Parmi de nombreuses autres questions, on a demandé aux hommes le nombre de travailleuses du sexe qu'ils avaient fréquentées au cours des douze derniers mois [11, Tableau 148-149]. On a demandé aussi aux travailleuses du sexe le nombre de clients qu'elles avaient eus pendant la semaine avant l'entretien [12, Tableau 78]. Le tableau 1 montre les résultats détaillés pour les 407 hommes du Yunnan. Les échantillons des différentes villes ont été pondérés selon la population de ces villes [11, Table 201] ; donc le tableau 1 ne montre pas des multiples entiers de la fraction 1/407. La première ligne du tableau 2 montre la moyenne et l'écart type, c'est-à-dire la racine carrée de la variance. Pour les 403 travailleuses du sexe du Yunnan qui ont participé à l'étude, seuls la moyenne, l'écart type, le minimum et le maximum ont été publiés (seconde ligne du tableau 2).

Table 1 – Distribution du nombre de travailleuses du sexe fréquentées au cours des douze derniers mois par 407 hommes du Yunnan. D'après [11, tableau 149].

	0	1	2	3	4	5	6-9	10-14	> 14
%	78,6	1,9	3,2	2,9	1,7	1,8	3,5	4,3	2,1

Table 2 – Distribution du nombre de travailleuses du sexe fréquentées au cours des douze derniers mois par 407 hommes et distribution du nombre de clients pendant une semaine pour 403 travailleuses du sexe dans le Yunnan. D'après [11, tableau 148] et [12, tableau 78].

	moyenne	écart type	minimum	maximum
hommes	1,0	2,2	0	non disponible
travailleuses	3,0	4,1	0	40

À Kunming, la capitale de la province du Yunnan, l'épidémie de VIH est devenue importante parmi les usagers de drogue en 1996-1997 (figure 1a). Depuis, il semblerait que la prévalence du VIH dans ce groupe à risque se soit stabilisée entre 20% et 30%, un niveau probablement déterminé par le pourcentage des usagers de drogue qui partagent les seringues. La transmission sexuelle du VIH semble assez limitée lorsqu'on la compare à ce qui est arrivé il y a quelques années dans d'autres régions de l'Asie du sud-est telles que la Thaïlande. La prévalence du VIH est restée relativement faible parmi les travailleuses du sexe et leurs clients (figure 1b, c). Noter que le petit accroissement après 1996 est probablement dû au fait que certaines travailleuses du sexe et certains clients sont aussi des usagers de drogue. Mais il n'y a pas d'indication d'une croissance

exponentielle de la transmission sexuelle. La situation dans d'autres parties du Yunnan a été résumée récemment [18].

L'absence de croissance exponentielle de la prévalence du VIH parmi les travailleuses du sexe suggère que la reproductivité nette R_0 pour la transmission sexuelle du VIH entre les travailleuses du sexe et leurs clients reste inférieure à 1. C'est un peu surprenant. Est-il possible de comprendre cela avec les données sur les comportements sexuels ?

Dans cet article, on utilise un modèle mathématique pour les épidémies de VIH qui prend en compte la distribution de l'activité sexuelle. Les premiers modèles de ce type se focalisaient sur les hommes homosexuels et divisaient cette population suivant le nombre k de partenaires sexuels par an ([1, 2, 15] et [4, chapitre 11]) : le compartiment k était appelé le « groupe d'activité sexuelle k », k étant un entier positif ou nul. Avec un mélange proportionnel, le seuil épidémique ne dépend pas seulement de la moyenne (notée M) mais aussi de la variance V de la distribution. Plus précisément, la reproductivité nette R_0 est proportionnelle à $M + V/M$ et l'épidémie ne se produit que si $R_0 > 1$. Pour une population homogène, la variance V est nulle et R_0 est simplement proportionnelle à l'activité sexuelle moyenne M. Ce modèle peut être modifié pour prendre en compte des distributions continues et non discrètes pour le taux d'acquisition de nouveaux partenaires sexuels [1, 2]. La reproductivité nette R_0 est alors toujours proportionnelle à $M + V/M$ [6, p. 81].

Le modèle a aussi été étendu aux populations hétérosexuelles ([15], [4, §11.3.9]) : si M et V sont la moyenne et la variance de l'activité sexuelle des hommes, si $\widehat{M}$ et $\widehat{V}$ sont la moyenne et la variance pour celle des femmes, alors la reproductivité nette R_0 est proportionnelle à

$$\sqrt{(M + V/M)(\widehat{M} + \widehat{V}/\widehat{M})}\,.$$

La variante avec des distributions continues et non discrètes pour le taux d'acquisition de nouveaux partenaires sexuels donne le même résultat pour R_0 [6, p. 83] (on peut remarquer dans cette référence que la formule pour R_0 est obtenue en considérant directement l'opérateur de prochaine génération, c'est-à-dire sans écrire explicitement les équations du modèle).

A priori, il semblerait que la moyenne et la variance du tableau 2 puissent être utilisées directement dans la formule pour R_0. Dans le présent article, on explique que ce n'est pas le cas. Supposons par exemple que le taux d'acquisition de nouveaux partenaires sexuels soit donné par une distribution continue. Les personnes dans le groupe d'activité sexuelle x ont un nouveau partenaire sexuel durant une durée infinitésimale dt avec une probabilité $x\,dt$ (x est un réel positif ou nul). Donc si l'on demande aux personnes dans ce groupe combien de nouveaux partenaires elles ont eus pendant une durée τ avant l'entretien, on s'attend à ce que la réponse suive une distribution de Poisson de moyenne $x\,\tau$. Ainsi, la distribution dans la population du nombre de partenaires sexuels sur une période donnée est une « loi mixte de Poisson » [8, chapitre 2] ; des distributions de ce type sont utilisées par les compagnies d'assurance pour modéliser le nombre de demandes d'une certaine catégorie, par exemple celles enregistrées

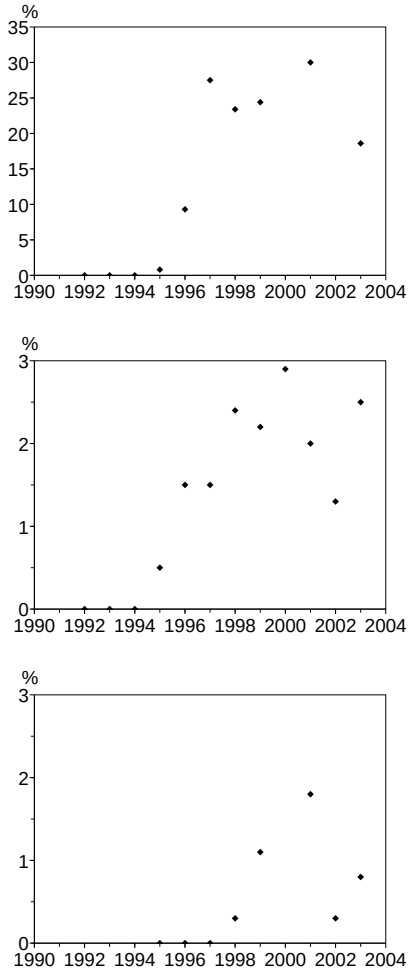

FIGURE 1 – En haut : prévalence du VIH parmi les utilisateurs de drogue à Kunming. Données de [19, 20, 21, 22]. En bas : prévalence du VIH parmi les travailleuses du sexe (à gauche) et parmi les clients (à droite). Données de [13]. Noter la différence d'échelles et voir [5] pour plus de références et de données.

après un accident de la circulation ou après une maladie, au cours d'une période donnée [8, chapitre 9]). Notons M et V la moyenne et la variance de la distribution de l'activité sexuelle, qui est une distribution de probabilité sur la demi-droite $x \geq 0$. Notons $m(\tau)$ et $v(\tau)$ la moyenne et la variance rapportées de la distribution du nombre de partenaires sexuels pendant une durée τ d'après une enquête comportementale ; noter que cette distribution de probabilité est sur l'ensemble des entiers positifs ou nuls. On peut montrer que $M = m(\tau)/\tau$ et $V = (v(\tau) - m(\tau))/\tau^2$, de sorte que

$$M + \frac{V}{M} = \frac{1}{\tau}\Big(m(\tau) + \frac{v(\tau)}{m(\tau)} - 1\Big). \tag{1}$$

Remarquer en particulier que $V \leq v(\tau)/\tau^2$. La variance dans l'enquête est toujours supérieure à la variance de la distribution de l'activité sexuelle.

Pour une population homogène, par exemple, une population ne comprenant d'un seul groupe d'activité sexuelle x, la distribution rapportée serait une distribution de Poisson ; dans ce cas, $v(\tau) = m(\tau)$. Alors le côté droit de (1) est égal à $m(\tau)/\tau$ et R_0 est simplement proportionnel à la moyenne rapportée du nombre de partenaires, comme on pourrait s'y attendre.

Si $\tau \to +\infty$, on a $1/\tau \to 0$ et

$$\frac{1}{\tau}\Big(m(\tau) + \frac{v(\tau)}{m(\tau)} - 1\Big) \simeq \frac{1}{\tau}\Big(m(\tau) + \frac{v(\tau)}{m(\tau)}\Big).$$

Donc le terme « -1 » dans (1) n'est important que quand τ est petit. Mais dans une enquête telle que celle faite dans le Yunnan, τ doit être petit parce que la personne que l'on interroge doit pouvoir se souvenir et compter tous ses partenaires sexuels pendant une durée τ. Cette question devient plus difficile à mesure que τ croît. Les concepteurs de l'enquête du Yunnan ont estimé que $\tau = 1$ an était une durée raisonnable pour les clients, de même que $\tau = 1$ semaine pour les travailleuses du sexe. C'est la méthode pratique pour résoudre une difficulté : estimer directement pour un client ou une travailleuse du sexe son groupe d'activité sexuelle x et construire la distribution d'activité sexuelle dans la population.

Enfin, on peut remarquer qu'une expression semblable à $m(\tau)+v(\tau)/m(\tau)-1$ du côté droit de l'équation (1) se trouve dans la théorie des graphes aléatoires lorsqu'on étudie la plus grande composante connexe ([16] et [6, p. 170]).

L'article est organisé de la manière suivante. Dans la section 2, on présente un modèle pour des hétérosexuels à base d'équations différentielles ordinaires (EDO), avec une activité sexuelle qui est une variable continue et qui tient compte du renouvellement des hommes et des travailleuses du sexe car la durée moyenne de « travail » dans le Yunnan est 2,5 années [12, tableau 63]). Comme les données pour le Yunnan contiennent deux unités de temps, l'année pour les hommes et la semaine pour les travailleuses du sexe, un modèle avec une structure continue pour l'activité sexuelle est bien plus adapté qu'un modèle avec une structure discrète. On fait aussi attention à l'équilibre nécessaire entre le nombre de relations sexuelles comptées du point de vue des hommes avec celui

du point de vue des travailleuses du sexe. On trouve la reproductivité nette R_0 pour le modèle par une méthode d'agrégation semblable à celle de [1, 2, 15] et de [4, chapitre 11]. Dans la section 3, on démontre la formule (1) et on l'utilise avec les données du tableau 2. De cette manière, on peut obtenir une borne supérieure pour la moyenne géometrique des probabilités de transmission dans le contexte de Kunming. Dans la section 4, on insiste sur la dépendance de la moyenne rapportée $m(\tau)$ et de la variance rapportée $v(\tau)$ par rapport à la durée τ sur laquelle les partenaires sexuels sont comptés. Notre analyse fait douter de la loi de puissance touvée dans [3], supposée lier $m(\tau)$ et $v(\tau)$ indépendamment de τ. La conclusion résume les objectifs de l'article.

2 Le modèle

Les compartiments du modèle utilisé dans cet article sont les suivants, la variable x étant réelle et positive :

- $S_0(t)$: nombre d'hommes susceptibles au temps t qui n'ont jamais de relation avec des travailleuses du sexe ;
- $S(t,x)$: densité de clients susceptibles au temps t tels que leur probabilité d'avoir un nouveau contact avec une travailleuse du sexe pendant une durée infinitésimale dt soit $x\,dt$; on dit que ces hommes appartiennent au « groupe d'activité sexuelle x » ;
- $I(t,x)$: densité de clients infectés dans le groupe d'activité sexuelle x ;
- $\widehat{S}(t,x)$: densité de travailleuses du sexe susceptibles dans le groupe d'activité sexuelle x, c'est-à-dire ayant une probabilité d'avoir un nouveau contact avec un client égale à $x\,dt$ pendant un intervalle de temps infinitésimal dt ;
- $\widehat{I}(t,x)$: densité de travailleuses du sexe infectées dans le groupe d'activité sexuelle x.

On utilise le terme « densité » parce que par exemple $\int_{x_1}^{x_2} S(t,x)\,dx$ est le nombre de clients susceptibles dont l'activité sexuelle est comprise entre x_1 et x_2. Notons

$$N(t,x) = S(t,x) + I(t,x)\,, \quad \widehat{N}(t,x) = \widehat{S}(t,x) + \widehat{I}(t,x)$$

la densité totale de clients et la densité totale de travailleuses du sexe. Les paramètres du modèle sont :

- A, $\widehat{A}(t)$: nombre de nouveaux hommes et de nouvelles travailleuses du sexe entrant dans la population adulte par unité de temps ; noter que l'arrivée de nouvelles travailleuses du sexe est supposée dépendre du temps ;
- ε $(0 < \varepsilon < 1)$: fraction d'hommes qui n'ont jamais de contact avec des travailleuses du sexe ;
- $F(x)$, $\widehat{F}(x)$: densité de probabilité selon laquelle les hommes et les travailleuses du sexe entrent dans le groupe d'activité sexuelle x ; on suppose que $\int_{0^+}^{\infty} F(x)\,dx = 1-\varepsilon$ et que $\int_0^{\infty} \widehat{F}(x)\,dx = 1$; la distribution d'activité sexuelle des hommes adultes est donc composée d'une masse de Dirac en $x = 0$ de masse ε, tandis que la fraction restante $1-\varepsilon$ est distribuée sur

la demi-droite $x > 0$; toutes les travailleuses du sexe sont actives ($x > 0$) par définition ;

- B, $\widehat{B}$: probabilité de transmission du VIH par relation des femmes aux hommes et des hommes aux femmes ; ce sont des moyennes qui ignorent l'hétérogénéité de la charge virale et la présence ou l'absence d'autres infections sexuellement transmissibles (ces deux aspects sont écartés dans le modèle parce que les données nécessaires ne sont pas disponibles) ; les personnes avec différents niveaux d'activité sexuelle pourraient aussi avoir différentes probabilités d'utiliser des préservatifs (B et $\widehat{B}$ seraient alors aussi des fonctions de x) mais il y a peu d'information à ce sujet dans le rapport concernant le Yunnan [11, tableau 184] ;
- C : taux auquel les hommes adultes quittent la population en question à cause du vieillissement ;
- $\widehat{C}$: taux auquel les travailleuses du sexe arrêtent leur travail ;
- $D = \widehat{D}$: taux auquel les personnes infectées développement le SIDA ; on suppose que les personnes avec le SIDA sortent de la population.

Introduisons quelques notations. Soit $G(x)$ la fonction de répartition de l'activité sexuelle chez les hommes :

$$G(x) = \begin{cases} 0 & \text{si} \quad x < 0, \\ \varepsilon + \int_0^x F(y)\,dy & \text{si} \quad x \geq 0. \end{cases}$$

Soit M et V la moyenne et la variance. Alors

$$M = \int_0^\infty x\,dG(x) = \int_0^\infty x\,F(x)\,dx, \tag{2}$$

$$V = \int_0^\infty x^2\,dG(x) - M^2 = \int_0^\infty x^2\,F(x)\,dx - M^2\,. \tag{3}$$

Noter que le paramètre ε n'apparaît pas explicitement dans ces formules mais qu'il est implicite puisque $\int_0^\infty F(x)\,dx = 1 - \varepsilon$. On définit de la même manière la moyenne $\widehat{M}$ et la variance $\widehat{V}$ pour l'activité sexuelle des travailleuses du sexe (sans avoir à discuter de $x = 0$). Les équations du modèle sont

$$\frac{dS_0}{dt}(t) = A\,\varepsilon - C\,S_0(t) \tag{4}$$

pour les hommes qui n'ont pas de contacts avec les travailleuses du sexe,

$$\frac{\partial S}{\partial t}(t,x) = A\,F(x) - C\,S(t,x) - B\,x\,S(t,x)\,\frac{\int_0^\infty y\,\widehat{I}(t,y)\,dy}{\int_0^\infty y\,\widehat{N}(t,y)\,dy}\,, \tag{5}$$

$$\frac{\partial I}{\partial t}(t,x) = B\,x\,S(t,x)\,\frac{\int_0^\infty y\,\widehat{I}(t,y)\,dy}{\int_0^\infty y\,\widehat{N}(t,y)\,dy} - (C+D)\,I(t,x) \tag{6}$$

pour les autres hommes, et

$$\frac{\partial \widehat{S}}{\partial t}(t,x) = \widehat{A}(t)\,\widehat{F}(x) - \widehat{C}\,\widehat{S}(t,x) - \widehat{B}\,x\,\widehat{S}(t,x)\,\frac{\int_0^\infty y\,I(t,y)\,dy}{\int_0^\infty y\,N(t,y)\,dy}\,, \tag{7}$$

$$\frac{\partial \widehat{I}}{\partial t}(t,x) = \widehat{B}\,x\,\widehat{S}(t,x)\,\frac{\int_0^\infty y\,I(t,y)\,dy}{\int_0^\infty y\,N(t,y)\,dy} - (\widehat{C}+\widehat{D})\,\widehat{I}(t,x)\,, \tag{8}$$

pour les travailleuses du sexe. On suppose qu'initialement

$$S_0(0) = A\,\varepsilon/C\,, \quad N(0,x) = A\,F(x)/C\,, \quad \widehat{N}(0,x) = A\,M\,\widehat{F}(x)/(C\,\widehat{M})\,, \tag{9}$$

$0 \le S(0,x) \le N(0,x)$ et $0 \le \widehat{S}(0,x) \le \widehat{N}(0,x)$ pour tout $x > 0$. Avec (4), on peut voir que $S_0(t) = A\,\varepsilon/C$ pour tout t. Comme le nombre de contacts sexuels comptés selon les clients devraient être égal au nombre de contacts sexuels comptés par les travailleuses du sexe, on doit avoir

$$\int_0^\infty x\,N(t,x)\,dx = \int_0^\infty x\,\widehat{N}(t,x)\,dx \tag{10}$$

pour tout t. Pour vérifier cette contrainte, on suppose quelque peu artificiellement que le recrutement des travailleuses du sexe $\widehat{A}(t)$ est donné par la formule

$$\begin{aligned}\widehat{A}(t)\,\widehat{M} = A\,M + (\widehat{C}-C)\int_0^\infty x\,S(t,x)\,dx + (\widehat{C}-C-D)\int_0^\infty x\,I(t,x)\,dx \\ + \widehat{D}\int_0^\infty x\,\widehat{I}(t,x)\,dx\,,\end{aligned} \tag{11}$$

ce qui signifie que les personnes qui recrutent peuvent équilibrer l'offre et la demande à chaque instant. En effet, il résulte des équations (5)-(8) et de (11) que

$$\begin{aligned}\frac{d}{dt}\int_0^\infty x\,\Big(N(t,x)-\widehat{N}(t,x)\Big)\,dx &= \int_0^\infty x\,\Big(\frac{\partial N}{\partial t}(t,x) - \frac{\partial \widehat{N}}{\partial t}(t,x)\Big)\,dx \\ &= A\,M - C\int_0^\infty x\,N(t,x)\,dx - D\int_0^\infty x\,I(t,x)\,dx \\ &\quad - \Big[\widehat{A}(t)\,\widehat{M} - \widehat{C}\int_0^\infty x\,\widehat{N}(t,x)\,dx - \widehat{D}\int_0^\infty x\,\widehat{I}(t,x)\,dx\Big] \\ &= -\widehat{C}\int_0^\infty x\Big(N(t,x) - \widehat{N}(t,x)\Big)\,dx\,.\end{aligned} \tag{12}$$

Vu le choix des conditions initiales (9), on a

$$\int_0^\infty x\Big(N(0,x) - \widehat{N}(0,x)\Big)\,dx = 0\,. \tag{13}$$

Il résulte de (12) et (13) que l'équilibre (10) est vérifié pour tout t.

Si l'on suppose aussi que $\widehat{C} > C + D$, alors le système (5)-(8) avec $\widehat{A}(t)$ défini par (11) est bien posé, $\widehat{A}(t)$ étant positif ou nul pour tout t. Ceci veut dire que la rotation des travailleuses du sexe doit être assez rapide pour s'adapter aux variations de l'offre et de la demande causées par la mortalité du SIDA. On verra dans la pochaine section que les données pour le Yunnan vérifient la condition $\widehat{C} > C + D$. Insistons sur les principales hypothèses du présent modèle :

- La durée des contacts n'est pas prise en compte. C'est l'hypothèse habituelle pour la modélisation des relations entre clients et travailleuses du sexe. Pour être plus précis, on considère les données sur la « fréquence moyenne des contacts avec une travailleuse du sexe au cours des 30 derniers jours » pour les hommes qui ont eu un contact au cours de l'année (soit 21,4% de tous les hommes d'après le tableau 1 ci-dessus) : la fréquence moyenne est 2,4 fois par mois [11, tableau 174]. Vu que le nombre moyen de travailleuses du sexe rencontrées par les hommes est 1,0 par an (tableau 2 ci-dessus), on peut estimer le nombre moyen de travailleuses du sexe rencontrées au cours d'une année (notons le μ) par les hommes pour lequel ce nombre est non nul, car $1{,}0 = 0 \times 78{,}6\% + \mu \times 21{,}4\%$. On obtient $\mu = 1{,}0/0{,}214 \simeq 4{,}67$. Le nombre moyen de contacts pour ces clients sur un an s'estime en multipliant le nombre mensuel par le nombre de mois dans une année : $2{,}4 \times 12 = 28{,}8$. Donc on obtient une moyenne de $28{,}8/4{,}67 \simeq 6{,}2$ contacts avec la même travailleuses du sexe et une durée moyenne du « contact » égale à $6{,}2/2{,}4 \simeq 2{,}6$ mois (avec une moyenne de 2,4 « liaisons » par an). Cette échelle de temps est beaucoup plus courte que par exemple la durée moyenne de « travail » $1/C'$ des travailleuses du sexe (2,5 années d'après [12, tableau 63]), donc l'hypothèse initiale est quelque peu justifiée.
- On suppose que le mélange entre les groupes d'activité sexuelle est proportionnel car on n'a pas de meilleure information dans l'enquête.
- À mesure que l'épidémie se développe, on suppose que le recrutement des travailleuses du sexe s'adapte pour équilibrer l'offre et la demande. Il y a plusieurs autres possibilités. Par exemple, les hommes pourraient fréquenter les travailleuses du sexe moins souvent. Ces hypothèses n'entrent de toute façon pas dans le calcul de la reproductivité nette R_0, qui est le point principal de cet article. Celui-ci ne dépend que du début de l'épidémie.

L'équilibre sans maladie du système (5)-(8) et (11) est donné par

$$S_*(x) = A\,F(x)/C, \quad \widehat{S}_*(x) = A\,M\,\widehat{F}(x)/(C\,\widehat{M}), \tag{14}$$

$I(x) = 0$, $\widehat{I}(x) = 0$ pour tout $x > 0$. Linéarisons les équations (6) et (8) près de cet équilibre sans maladie et utilisons des lettres minuscules pour éviter toute

confusion. On obtient

$$\frac{\partial i}{\partial t}(t,x) = B\,x\,S_*(x)\,\frac{\int_0^\infty y\,\widehat{i}(t,y)\,dy}{\int_0^\infty y\,\widehat{S}_*(y)\,dy} - (C+D)\,i(t,x) \tag{15}$$

$$\frac{\partial \widehat{i}}{\partial t}(t,x) = \widehat{B}\,x\,\widehat{S}_*(x)\,\frac{\int_0^\infty y\,i(t,y)\,dy}{\int_0^\infty y\,S_*(y)\,dy} - (\widehat{C}+\widehat{D})\,\widehat{i}(t,x)\,. \tag{16}$$

Suivant la même méthode que celle de [4, §11.3.9] pour un modèle avec une distribution discrète de l'activité sexuelle, on introduit les variables agrégées

$$J(t) = \int_0^\infty x\,i(t,x)\,dx, \quad \widehat{J}(t) = \int_0^\infty x\,\widehat{i}(t,x)\,dx\,.$$

Il résulte de (15)-(16) que

$$\frac{dJ}{dt}(t) = B\frac{\int_0^\infty x^2\,S_*(x)\,dx}{\int_0^\infty x\,\widehat{S}_*(x)\,dx}\,\widehat{J}(t) - (C+D)\,J(t)$$

$$\frac{d\widehat{J}}{dt}(t) = \widehat{B}\frac{\int_0^\infty x^2\,\widehat{S}_*(x)\,dx}{\int_0^\infty x\,S_*(x)\,dx}\,J(t) - (\widehat{C}+\widehat{D})\,\widehat{J}(t)\,.$$

Avec (2)-(3) et (14), ce système peut s'écrire

$$\frac{dJ}{dt}(t) = B\,\frac{V+M^2}{M}\,\widehat{J}(t) - (C+D)\,J(t) \tag{17}$$

$$\frac{d\widehat{J}}{dt}(t) = \widehat{B}\,\frac{\widehat{V}+\widehat{M}^2}{\widehat{M}}\,J(t) - (\widehat{C}+\widehat{D})\,\widehat{J}(t)\,. \tag{18}$$

On peut alors facilement montrer que l'équilibre nul de ce dernier système est linéairement stable si et seulement si

$$R_0 = \sqrt{\frac{B\,\widehat{B}(M+V/M)(\widehat{M}+\widehat{V}/\widehat{M})}{(C+D)\,(\widehat{C}+\widehat{D})}} < 1\,. \tag{19}$$

Commentons cette formule. On aurait pu arriver au même résultat en n'incluant pas dans le modèle les hommes qui n'ont jamais de contact avec les travailleuses du sexe, c'est-à-dire en restreignant notre attention aux clients des travailleuses du sexe. En effet, la distribution de l'activité sexuelle pour ces clients est $\widetilde{F}(x) = F(x)/(1-\varepsilon)$ et l'activité sexuelle moyenne est $\widetilde{M} = M/(1-\varepsilon)$. Noter que $\widetilde{M} > M$. Mais si $\widetilde{V}$ est la variance correspondante, alors

$$M + \frac{V}{M} = \frac{\int_0^\infty x^2 F(x)\,dx}{\int_0^\infty x\,F(x)\,dx} = \frac{\int_0^\infty x^2\widetilde{F}(x)\,dx}{\int_0^\infty x\,\widetilde{F}(x)\,dx} = \widetilde{M} + \frac{\widetilde{V}}{\widetilde{M}}\,.$$

Donc on obtient exactement le même R_0.

3 Lien avec les données

Les personnes dans le groupe d'activité sexuelle x ont un nombre de partenaires sur une période de durée τ qui est distribué selon une loi de Poisson de moyenne $x\,\tau$ [7, §XVII.2]. Donc leur probabilité d'avoir j partenaires sur une période de durée τ est

$$e^{-x\tau}\,(x\tau)^j/j!$$

Par conséquent, la fraction des hommes qui déclarent j partenaires sur une période de durée τ dans une enquête devrait être

$$f_j(\tau) = \int_0^\infty e^{-x\tau}\,\frac{(x\tau)^j}{j!}\,dG(x) = \begin{cases} \varepsilon + \int_{0^+}^\infty e^{-x\tau}\,F(x)\,dx & \text{si} \quad j=0, \\ \int_0^\infty e^{-x\tau}\,\frac{(x\tau)^j}{j!}\,F(x)\,dx & \text{si} \quad j\geq 1. \end{cases} \tag{20}$$

Autrement dit, $(f_j(\tau))_{j\geq 0}$ est une loi mixte de Poisson [8, chapitre 2]. La moyenne rapportée $m(\tau)$ et la variance $v(\tau)$ sont données par

$$m(\tau) = \sum_{j\geq 1} j\,f_j(\tau)\,, \quad v(\tau) = \sum_{j\geq 1} j^2\,f_j(\tau) - m(\tau)^2\,. \tag{21}$$

Il résulte de (2), (20) et (21) que la moyenne rapportée $m(\tau)$ est égale à τM (voir aussi [8, proposition 2.1(i)]) :

$$m(\tau) = \int_0^\infty e^{-x\tau} \sum_{j\geq 1} \frac{(x\tau)^j}{(j-1)!}\,dG(x) = \tau \int_0^\infty x\,dG(x) = \tau M\,. \tag{22}$$

Par ailleurs, il résulte de (20) que

$$\sum_{j\geq 1} j(j-1)f_j(\tau) = \int_0^\infty e^{-x\tau} \sum_{j\geq 2} \frac{(x\tau)^j}{(j-2)!}\,dG(x) = \tau^2 \int_0^\infty x^2\,dG(x)\,. \tag{23}$$

Combinons l'expression (3) de la variance V avec (21)-(22)-(23). On voit que

$$\begin{aligned} V &= \frac{1}{\tau^2}\sum_{j\geq 1} j(j-1)f_j(\tau) - \Big(\frac{m(\tau)}{\tau}\Big)^2 \\ &= \frac{1}{\tau^2}\sum_{j\geq 1} j^2\,f_j(\tau) - \frac{1}{\tau^2}\sum_{j\geq 1} j\,f_j(\tau) - \frac{m(\tau)^2}{\tau^2} = \frac{v(\tau)-m(\tau)}{\tau^2} \end{aligned} \tag{24}$$

(voir aussi [8, proposition 2.1(ii)]). Noter ici que puisque $m(\tau) \geq 0$ et $V \geq 0$, toute loi mixte de Poisson a deux propriétés importantes : $v(\tau) \geq m(\tau)$ (l'inégalité est stricte si $V > 0$) et $v(\tau) \geq \tau^2\,V$ (l'inégalité est stricte si $m(\tau) > 0$, c'est-à-dire si $M > 0$). Enfin, (22) et (24) donnent

$$M + \frac{V}{M} = \frac{m(\tau)}{\tau} + \frac{(v(\tau)-m(\tau))/\tau^2}{m(\tau)/\tau} = \frac{1}{\tau}\Big(m(\tau) + \frac{v(\tau)}{m(\tau)} - 1\Big)\,. \tag{25}$$

De même, on obtient pour les travailleuses du sexe sur une période τ' la formule

$$\widehat{M} + \frac{\widehat{V}}{\widehat{M}} = \frac{1}{\widehat{\tau}}\Big(\widehat{m}(\widehat{\tau}) + \frac{\widehat{v}(\widehat{\tau})}{\widehat{m}(\widehat{\tau})} - 1\Big)\,. \tag{26}$$

Remarques :

- Avec les expressions (2)-(3) pour M et V, on voit que la dimension physique (ou unité) de x est [temps]$^{-1}$, celle de M est [temps]$^{-1}$, celle de V est [temps]$^{-2}$, celle de τ est [temps], mais m et v sont sans dimension (ce sont des nombres rapportés). On peut vérifier que les deux côtés de la formule (25) ont la même dimension.
- Si $\widehat{F}(x)$ est une distribution Gamma comme dans [1, 2], alors on peut montrer que la distribution discrète associée $(\widehat{f}_j(\widehat{\tau}))_{j\geq 0}$, donnée par

$$\widehat{f}_j(\widehat{\tau}) = \int_0^\infty e^{-x\widehat{\tau}}\,\frac{(x\widehat{\tau})^j}{j!}\,\widehat{F}(x)\,dx\,,$$

  est une loi binomiale négative [8, p. 17]. Ce point, déjà noté dans [10], ne semble pas avoir été remarqué dans la littérature sur les modèles d'activité sexuelle pour le VIH/SIDA.

On souhaiterait maintenant appliquer les formules (19), (25) et (26) aux données du tableau 2 pour obtenir une estimation de R_0. Le problème est que certains paramètres ne sont pas bien connus. D'un côté, on connaît le taux $\widehat{C}$ auquel les travailleuses du sexe arrêtent leur travail : puisque la durée moyenne de ce travail est de 2,5 années [12, tableau 63], on peut prendre $\widehat{C} = 1/2{,}5 = 0{,}4$ par an. Le pourcentage d'hommes qui ont eu un contact avec une travailleuse du sexe au cours de l'année est relativement constant en fonction de l'âge (tableau 3). Donc pour le paramètre C, on garde le taux moyen de vieillissement entre 18 et 50 ans, c'est-à-dire $C = 1/(50-18) = 1/32$ par an. Pour le taux de progression du VIH au SIDA, on prend pour simplifier une constante $D = \widehat{D} = 1/10$ par an, ce qui donne une incubation moyenne de 10 ans. Noter que $\widehat{C} > C + D$, comme requis dans la section précédente.

Table 3 – Pourcentage des hommes ayant eu un contact avec une travailleuse du sexe au cours des 12 derniers mois [11, tableau 173].

âge	18-30	30-34	35-39	40-44	45-50
avec contact	25,6%	25,2%	20,2%	19,1%	8,8%

Des problèmes proviennent cependant de l'incertitude concernant les probabilités de transmission par relation B et $\widehat{B}$. On a estimé dans la section 2 qu'une relation dans le présent contexte signifie en moyenne 6 contacts sexuels. Une étude de la transmission du VIH entre les travailleuses du sexe et leurs clients en Thaïlande [14] suggère que la probabilité de transmission par contact sexuel de la femme à l'homme peut atteindre 3% à cause de la prévalence élevée des autres infections sexuellement transmissibles chez les travailleuses du sexe. D'autres études ont trouvé des probabilités de transmission par contact bien

moindres, par exemple 0,11% pour des couples sérodiscordants en Ouganda [9]. Il est difficile de dire quelle estimation entre ces deux valeurs extrêmes conviendrait pour Kunming. En plus, on remarque que l'utilisation de préservatifs est incluse sous forme moyennée dans les paramètres B et $\widehat{B}$.

Étant données toutes ces difficultés, on retourne la question de la manière suivante : sachant que l'on ne voit pas encore de croissance exponentielle de la transmission sexuelle du VIH, que peut-on en déduire pour les paramètres inconnus du modèle ? Avec $R_0 < 1$ et avec (19), (25) et (26), on obtient une borne supérieure pour la moyenne géometrique de ces probabilités de transmission moyennées par relation :

$$\begin{aligned}\sqrt{B\,\widehat{B}} &< \sqrt{\frac{\tau\,\widehat{\tau}\,(C+D)\,(\widehat{C}+\widehat{D})}{(m+\frac{v}{m}-1)(\widehat{m}+\frac{\widehat{v}}{\widehat{m}}-1)}} \\ &= \sqrt{\frac{1\times(7/365)\times(0{,}4+0{,}1)\times(1/32+0{,}1)}{(1{,}0+\frac{2{,}2^2}{1{,}0}-1)(3{,}0+\frac{4{,}1^2}{3{,}0}-1)}} \simeq 0{,}58\%\,.\end{aligned}$$

Avec l'hypothèse simplificatrice supplémentaire $B = \widehat{B}$ et en considérant (voir la section 2) qu'une relation représente en moyenne 6 contacts sexuels, on peut obtenir une borne supérieure pour la probabilité de transmission par contact sexuel (notons la b) : $b < 0{,}1\%$. En effet, $1-(1-0{,}1\%)^6 \simeq 0{,}6\%$. Donc notre résultat est plus proche de l'estimation inférieure pour la probabilité de transmission par contact sexuel, mais ceci est peut-être dû au niveau élevé d'utilisation des préservatifs. Dans l'enquête comportementale de 2002 au Yunnan, 73,5% des clients ont déclaré avoir utilisé un préservatif lors de leur dernier rapport sexuel avec une travailleuse du sexe [11, tableau 177].

Si l'on observait une croissance exponentielle de la transmission sexuelle, on pourrait estimer la moyenne géométrique $\sqrt{B\,\widehat{B}}$ puisque le taux de croissance λ du modèle est la plus grande valeur propre de la matrice du côté droit du système (17)-(18)

$$\lambda = \frac{-(C+\widehat{C}+D+\widehat{D}) + \sqrt{(C-\widehat{C})^2 + 4B\widehat{B}(M+V/M)(\widehat{M}+\widehat{V}/\widehat{M})}}{2}\,.$$

Elle est liée au temps de doublement T par $T = \log 2/\lambda$. Le point principal ici est que M et V ne peuvent être lus directement sur les données ; on doit utiliser $m(\tau)$ et $v(\tau)$ avec le « -1 » additionnel dans les formules (25)-(26).

4 Lien entre la moyenne et la variance

Dans [3], on peut lire :

> « *a wide range of surveys of different populations employing different sampling methods and various time intervals for recall, reveal a remarkably consistent trend in the relationship between the mean and*

variance in the rate of acquisition of new partners. The summary statistics are related by a power law, $v = a\,m^b$, where a and b are constants. »

Les constantes obtenues étaient $a = 0{,}555$ et $b = 3{,}231$. On remarque que dans le cadre de notre modèle, cette loi de puissance ne peut être vraie pour « *various time intervals for recall [...] ranging from the past month to lifetime* ». En effet, l'intervalle de temps est τ selon nos notations. Puisque les équations (22) et (24) impliquent que $m(\tau) = \tau\,M$ et $v(\tau) = \tau^2\,V + \tau\,M$, et puisque $v(\tau)$ n'est pas une fonction homogène de τ (il n'y a pas de α tel que $v(s\tau) = s^\alpha\,v(\tau)$ pour tout $s > 0$), la loi de Poisson mixte du présent modèle semble incompatible avec la loi de puissance de [3]. Autrement dit, même si le point representant la moyenne et la variance rapportées sur une période fixée se trouve proche de la courbe pour la loi de puissance (comme c'est le cas pour les travailleuses du sexe dans le Yunnan, voir la figure 2), le point représentant la moyenne et la variance pour la même population mais pour une durée différente peut en être très éloigné. C'est ce qu'illustre la figure 2. Voir [3, figure 2a] pour le nuage de points original, qui devait justifier la loi de puissance. Pour faciliter la comparaison, l'échelle de la figure 2 est la même que celle de [3, figure 2a]. Avec le modèle utilisé dans le présent article, on remarque que $m(\tau) = \tau\,M$ et $v(\tau) = \tau^2\,V + \tau\,M$ impliquent que $v(\tau)/m(\tau)^2$ est à peu près constant (c'est-à-dire indépendant de τ) pour de grandes valeurs de τ et que $v(\tau)/m(\tau)$ est à peu près constant pour de petites valeurs de τ.

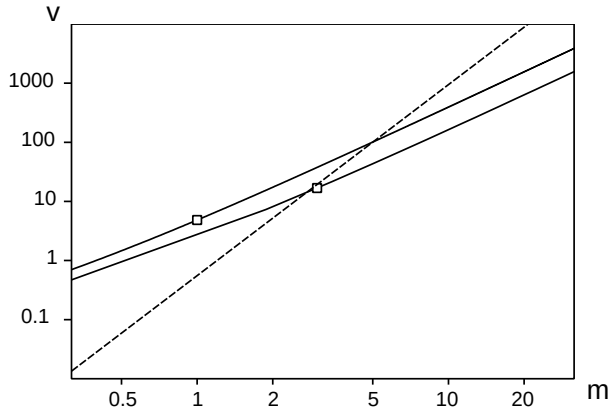

FIGURE 2 – Variance rapportée $v(\tau)$ en fonction du nombre moyen rapporté de partenaires $m(\tau)$ (échelle logarithmique). Lignes continues : ($m(\tau) = \tau\,M$, $v(\tau) = \tau^2\,V + \tau\,M$) avec (M, V) fixé et un intervalle τ variable. Points : données du tableau 2 (hommes à gauche, travailleuses du sexe à droite). En pointillé : loi de puissance $v = a\,m^b$ de [3].

5 Conclusion

L'influence des modèles mathématiques a été analysée dans [17] de la manière suivante [1] :

> *The results of several modeling efforts, especially those of Roy Anderson and colleagues [...] have shown that the rate of partner change is one of the key factors influencing the speed and size of the epidemic [...] Although there is little evidence that this understanding has influenced program design to any great extent, it has certainly influenced research and evaluation efforts. Several of the key prevention indicators developed by GPA [2], UNAIDS, and USAID are designed to measure rates of partner change and concurrent partnerships. If these indicators are seriously applied, they will eventually influence program decisions by showing which interventions improve these indicators and which do not affect them.*

Dans cet article, on a essayé d'appliquer « sérieusement » l'indicateur $M+V/M$ aux données réelles du Yunnan. On a trouvé des relations simples entre les paramètres théoriques (M, V) d'un côté, la moyenne $m(\tau)$ et la variance $v(\tau)$ rapportées pour une période de temps τ d'un autre côté. Il s'avère que l'expression pour l'indicateur $M + V/M$ en fonction de $m(\tau)$ et de $v(\tau)$ a une forme semblable à un paramètre important de la théorie des graphes aléatoires. Notre formulation fait aussi douter de la mystérieuse loi de puissance de [3].

Remerciements. Les auteurs ont reçu le soutien du « Programme de Recherches Avancées de Coopérations Franco-Chinoises » (PRA SI05-01, « Modélisation de l'épidémie de VIH/SIDA dans les provinces du Yunnan et du Xinjiang »).

Références

[1] R. M. Anderson and R. M. May, *The invasion, persistence and spread of infectious diseases within animal and plant communities*, Philos. Trans. R. Soc. Lond. B, **314** (1986), 533–570.

[2] R. M. Anderson, G. F. Medley, R. M. May, and A. M. Johnson, *A preliminary study of the transmission dynamics of the human immunodeficiency virus (HIV), the causative agent of AIDS*, IMA J. Math. Appl. Med. Biol., **3** (1986), 229–263.

[3] R. M. Anderson and R. M. May, *Epidemiological parameters of HIV transmission*, Nature, **333** (1988), 514–519.

1. Noter que l'auteur de [17] est vice-président de *Futures Group*. Ceci peut expliquer pourquoi l'enquête menée au Yunnan et au Sichuan par *Futures Group Europe* rapporte à la fois la moyenne et la variance pour la distribution du nombre de partenaires sexuels.
2. GPA : *Global Program on AIDS*.

[4] R. M. Anderson and R. M. May, *Infectious Diseases of Humans : Dynamics and Control*, Oxford University Press, 1991.

[5] N. Bacaër, X. Abdurahman, and J. Ye, *Modeling the HIV/AIDS epidemic among injecting drug users and sex workers in Kunming, China*, Bull. Math. Biol., **68** (2006), 525–550.

[6] O. Diekmann and J. A. P. Heesterbeek, *Mathematical Epidemiology of Infectious Diseases*, John Wiley & Sons, New York, 2000.

[7] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. I, 3e éd., John Wiley & Sons, New York, 1968.

[8] J. Grandell, *Mixed Poisson Processes*, Chapman & Hall, London, 1997.

[9] R. H. Gray, M. J. Wawer, R. Brookmeyer et al., *Probability of HIV-1 transmission per coital act in monogamous, heterosexual, HIV-1-discordant couples in Rakai, Uganda*, Lancet, **357** (2001), 1149-1153.

[10] M. Greenwood and G. U. Yule, *An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurence of multiple attacks of disease or of repeated accidents*, J. Roy. Statist. Soc., **83** (1920), 255–279.

[11] *Horizon Market Research and Futures Group Europe, 2001 Behavioural Surveillance Survey in Yunnan and Sichuan : Adult Male Report*, Décembre 2002, `http://www.futuresgroup.com/Documents/2001BSSadultmale.pdf`.

[12] *Horizon Market Research and Futures Group Europe, 2001 Behavioural Surveillance Survey in Yunnan and Sichuan : Sex Worker Report*, Décembre 2002, `http://www.futuresgroup.com/Documents/2001BSSsexworker.pdf`.

[13] L. Lu, M. Jia, X. Zhang et al., *Analysis for epidemic trend of acquired immunodeficiency syndrome in Yunnan Province of China*, Chin. J. Prev. Med., **38** (2004), 309–312.

[14] T. D. Mastro, G. A. Satten, T. Nopkesorn et al., *Probability of female-to-male transmission of HIV-1 in Thailand*, Lancet, **434** (1994), 204–207.

[15] R. M. May and R. M. Anderson, *Transmission dynamics of HIV infection*, Nature, **326** (1987), 137–142.

[16] M. Molloy and B. Reed, *The size of the giant component of a random graph with a given degree sequence*, Combin. Probab. Comput., **7** (1998), 295–305.

[17] J. M. Stover, *Influence of mathematical modeling of HIV and AIDS on policies and programs in the developing world*, Sex. Transm. Dis., **27** (2000), 572–578.

[18] Y. Xiao, S. Kristensen, J. Sun et al., *Expansion of HIV/AIDS in China : Lessons from Yunnan Province*, Soc. Sci. Med., **64** (2007), 665–675.

[19] H. Yu, X. An, M. Jia et al., *Report on HIV/AIDS surveillance in Yunnan province in 1999*, J. Chin. AIDS/STD Prev. Cont., **7** (2001), 74–76.

[20] J. Zhang, H. Cheng, M. Jia et al., *Ten years of experience on AIDS control in Yunnan (1989∼1998)*, Chin. J. Epidemiol., **20** (1999), 377–380.

[21] X. Zhang, Y. Ma, H. Yu et al., *Analysis on the surveillance result of HIV/AIDS in Yunnan in 2001*, Ji Bing Jian Ce, **17** (2002), 327–330.

[22] X. Zhang, J. Lu, L. Fu et al., *Analysis on survey results of HIV/AIDS in Yunnan Province in 2003*, Ji Bing Jian Ce, **19** (2004), 409–412.