

---

Vers un modèle prévisionnel de développement de l'infection VIH en France à partir de l'enquête ACSF

Author(s): Françoise Le Pont

Reviewed work(s):

Source: *Population (French Edition)*, 48e Année, No. 5, Sexualité et sciences sociales: Les apports d'une enquête (Sep. - Oct., 1993), pp. 1535-1550

Published by: [Institut National d'Études Démographiques](#)

Stable URL: <http://www.jstor.org/stable/1534188>

Accessed: 15/10/2012 03:19

---

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



*Institut National d'Études Démographiques* is collaborating with JSTOR to digitize, preserve and extend access to *Population (French Edition)*.

# VERS UN MODÈLE PRÉVISIONNEL DE DÉVELOPPEMENT DE L'INFECTION VIH EN FRANCE À PARTIR DE L'ENQUÊTE ACSF

Françoise LE PONT

*Le virus du Sida se transmet à l'occasion de contacts sexuels et par des échanges de seringues entre toxicomanes. Que peut-on dire sur l'évolution future de l'épidémie à partir des connaissances disponibles aussi bien sur les comportements à risque que sur l'épidémiologie de cette infection ? Quelles peuvent-être les conséquences de changements de comportement à la suite de campagnes de prévention ? Les modèles de diffusion du VIH peuvent apporter des réponses et permettre de tester des hypothèses aussi bien biologiques que sociologiques. Afin de prendre en compte le réseau de relations, l'évolution du comportement sexuel d'un individu au cours du temps et la variabilité du comportement d'un individu à l'autre, une méthode de simulation individuelle a été mise au point. Il faut pour cela s'appuyer sur les données d'une enquête, comme l'enquête ACSF, qui permette de construire une sorte d'instantané du réseau de relations sexuelles, et de dériver les règles de comportement nécessaires au fonctionnement du modèle. Françoise LE PONT\* a entrepris ce travail de simulation dont elle présente ici les grandes lignes avec quelques résultats préliminaires.*

Selon les données du 30 juin 1992 (WHO, 1992), la France est le pays d'Europe qui présente le nombre le plus élevé de cas de Sida et le plus fort taux de cas par million d'habitants après la Suisse, que l'on considère tous les groupes de transmission ou uniquement les cas d'origine hétérosexuelle. Le taux de cas de Sida d'origine transfusionnelle est, de loin, le plus élevé d'Europe (WHO, 1992). Cependant, les 21 000 cas de Sida déclarés en France, dont 12 % d'origine hétérosexuelle, ne représentent qu'une proportion, inconnue, de la population infectée par le VIH, en raison de la sous-déclaration et du délai de déclaration des cas de Sida, et surtout

---

\* INSERM U 263.

en raison du long délai d'incubation qui sépare la contamination de l'entrée dans le stade Sida (50 % des individus infectés développent le Sida en plus de 10 ans).

Des techniques ont donc été développées pour étudier et prévoir la croissance de l'épidémie. Hormis celles qui sont basées sur le nombre de cas de Sida et qui permettent des prévisions à court terme (extrapolation de la courbe des cas de Sida, ou estimation du nombre annuel d'individus contaminés dans le passé à partir de l'évolution du nombre annuel de cas de Sida et de la distribution de la période d'incubation), un grand nombre de *modèles explicatifs* ont été développés (Le Pont, 1989). Ces modèles intègrent les connaissances disponibles en socio-démographie, épidémiologie et biologie. Ils peuvent être considérés comme des outils exploratoires quand les observations directes sont difficiles ou impossibles à réaliser, en raison de problèmes pratiques, financiers, éthiques ou de délai.

Ainsi, les modèles explicatifs peuvent-ils apporter une meilleure compréhension de l'épidémie observée. On peut, par exemple, se demander pourquoi l'épidémie observée à Amsterdam chez les hommes homo/bisexuels depuis 1980 présente une dynamique si différente de celle observée à San Francisco. Ou encore, pourquoi la France est beaucoup plus touchée que la Grande-Bretagne alors que la période d'introduction du VIH dans les deux pays est probablement proche (environ 5 % des cas de Sida ont été rapportés avant 1986 et près de 40 % avant 1989 (WHO, 1992)). Des hypothèses peuvent être testées avec les modèles explicatifs par exemple celles qui concernent la prévalence et l'évolution temporelle des comportements à risque dans ces populations.

Les modèles explicatifs permettent également d'estimer certains paramètres de transmission de l'infection, d'orienter la collecte de données vers les facteurs qui ont une forte influence sur la dynamique de l'épidémie, de tester l'efficacité des politiques de prévention et de faire des prévisions. L'estimation de ces paramètres par modélisation est particulièrement utile lorsque les mesures sont difficiles, longues, voire impossibles à obtenir par des enquêtes épidémiologiques. Par exemple, la mesure du risque d'être contaminé par transfusion sanguine est réalisée en recherchant la proportion de receveurs de transfusion qui ont été contaminés : ces enquêtes, souvent difficiles et longues, ne sont pas exhaustives, ce qui induit un grand intervalle de confiance du risque estimé. En revanche, la simulation de la collecte des dons de sang a permis de mettre en évidence la forte diminution du risque transfusionnel entre 1988 et 1992 en France (Le Pont, 1993c). Pour sélectionner le meilleur estimateur d'un paramètre, les qualités de ces estimateurs peuvent être mesurées en simulant une population fictive et en déterminant si les valeurs obtenues concordent avec les caractéristiques de la population simulée.

En déterminant les paramètres qui ont une forte influence sur la croissance de l'épidémie (c'est-à-dire pour lesquels de faibles variations indui-

sent de fortes variations de l'épidémie simulée), les modèles servent à orienter le recueil des données. Par exemple, si l'on montre que la structure des réseaux relationnels est un facteur déterminant de la dynamique de l'épidémie, il sera alors souhaitable de mettre au point des enquêtes sur ces réseaux dans la population (voir à ce propos l'article de Ferrand et Mounier, dans le même numéro).

Les modèles permettent également de tester des scénarios afin, par exemple, d'étudier si l'instauration du dépistage de l'antigène p24 dans les dons de sang en plus du dépistage des anticorps permettrait de réduire le risque transfusionnel et quel serait le coût d'une contamination évitée (Le Pont, 1993c), ou si la fréquence d'utilisation des préservatifs déclarée chez les 18-24 ans annule le risque d'épidémie dans ce groupe (Le Pont, 1993a).

Enfin, lorsque les données adéquates sont disponibles, ces techniques peuvent permettre de faire des prédictions, par exemple de la taille de l'épidémie post-transfusionnelle en France (Le Pont, 1990) ou du risque d'une épidémie due à la transmission hétérosexuelle dans la population générale française (Le Pont, 1993a).

## I. – Les modèles de diffusion de l'épidémie

### ***Modèles «compartimentaux» et modèles de simulation***

Les modèles de la théorie des épidémies peuvent être classés, selon les méthodes utilisées, en *modèles mathématiques compartimentaux* (déterministes ou stochastiques, décrits par des équations mathématiques) et de *simulation* (Bailey, 1975).

Dans le cas des modèles mathématiques compartimentaux, la population est divisée en sous-populations appelées compartiments, dans lesquels tous les individus sont identiques (même classe d'âge, même état de santé, même sexe...). Ces modèles sont basés sur la loi d'action de masse introduite par Ross en 1911, suite aux travaux de Bernoulli en 1760, qui indique que le nombre de nouvelles infections est proportionnel à la fois au nombre de personnes sensibles et au nombre de personnes contagieuses. Dans le cas des modèles mathématiques *déterministes*, pour une situation donnée au départ, une solution unique est obtenue car les variations aléatoires ne sont pas prises en compte. Ceci est justifié lorsque les populations sensible et infectée sont grandes, lorsque les individus infectés ont beaucoup de contacts pendant les périodes contagieuses et lorsque les contacts entre les individus des différentes sous-populations se font au hasard. Mais ces hypothèses sont rarement vérifiées : par exemple, l'hypothèse de contact au hasard entre les individus est peu réaliste pour de grandes populations. Cette approche a défini le concept du taux de reproduction de base  $R_0$ , qui est un indicateur classiquement utilisé pour prédire s'il y aura ou non une épidémie dans une population (Bailey, 1975). A la différence des modèles

déterministes, les modèles *stochastiques* prennent en compte la variabilité due au hasard, qui est particulièrement importante dans le cas des petites populations. Dans ces modèles, les nombres d'individus sensibles et infectés sont considérés comme des variables aléatoires. Les modèles stochastiques sont souvent plus difficiles à résoudre mathématiquement que les modèles déterministes, et le résultat moyen obtenu avec un modèle stochastique n'est pas forcément identique au résultat obtenu avec un modèle déterministe. Cependant, la différence entre les résultats obtenus tend à s'estomper lorsque la taille de la population augmente (Bailey, 1975).

Alors que les modèles mathématiques traitent des populations, les modèles de simulation peuvent utiliser une *représentation individuelle*. Ils ont en particulier été utilisés en démographie et en économie pour étudier par exemple la reproduction (voir les références dans Leridon, 1977) ou la situation familiale et financière des futures retraitées (Pennec, 1993). A l'inverse de la plupart de ces modèles, ceux appliqués à la simulation des épidémies peuvent prendre en compte les interactions entre les individus. Ces modèles sont basés sur la méthode de Monte-Carlo (ou tirage aléatoire dans des lois de distribution) pour prendre en compte la variabilité des processus et des phénomènes biologiques, la population pouvant être examinée à intervalles de temps réguliers pour faire évoluer l'état du système. La simulation événementielle consiste à simuler les événements relatifs à l'évolution du système, ces événements étant rangés, au cours de la simulation, par date d'occurrence dans un agenda appelé échancier. L'inconvénient de ce type de modèle est la durée des simulations et la difficulté d'extraire des règles générales à partir des résultats obtenus, car on ne dispose pas toujours de méthodes permettant d'explorer l'ensemble des situations possibles.

Parmi la cinquantaine de modèles de diffusion du VIH et du Sida présentés dans la littérature, la très grande majorité sont des modèles mathématiques compartimentaux déterministes (Le Pont, 1989). Les modèles mathématiques ont permis de mettre en évidence des facteurs sociologiques qui influent sur la diffusion de l'épidémie. Ainsi, à partir des données concernant des populations d'Afrique sub-saharienne (Anderson, 1991), l'impact démographique du Sida prévu par les modèles est plus important si les hommes ont des rapports sexuels avec des femmes plus jeunes qu'avec des femmes de même âge. Des travaux portant sur des populations d'hétérosexuels toxicomanes ou non toxicomanes du Bronx aux USA (Blower, 1991) ont montré qu'un ralentissement de la croissance de l'épidémie peut être transitoire et ne signifie pas que les individus ont modifié leur comportement; il peut provenir de la diffusion du virus de la population toxicomane à haut risque vers la population moins exposée non toxicomane. D'autres travaux ont également démontré que les échanges entre les groupes de la population sont des facteurs déterminants de la diffusion du VIH, par exemple au sein d'une population d'hommes homosexuels américains (Jacquez, 1989). Dans cette étude, la population est divisée en sous-groupes selon le taux annuel de rencontre de partenaires. Les échanges

intra- et inter-groupes sont définis par des matrices de choix du partenaire qui indiquent les proportions de partenaires qu'un individu d'un groupe donné choisit dans son groupe et dans les autres groupes. Selon les valeurs indiquées dans la matrice, on définit une population dans laquelle les échanges se font de préférence avec des individus qui se ressemblent («like-with-like mixing») ou qui sont différents («assortative mixing»). Les simulations montrent que la phase de croissance initiale de l'épidémie dépend des échanges intra-groupes pour les groupes à fort taux de rencontre de partenaire, alors que la taille de l'épidémie dépend des contacts inter-groupes. Si le taux de rencontre intra-groupe est élevé, l'épidémie peut présenter plusieurs pics, ce qui rejoint les résultats présentés dans la population du Bronx (Blower, 1991).

### ***Hétérogénéité et niveaux d'analyse***

On voit donc qu'un modèle de diffusion de l'épidémie doit pouvoir prendre en compte l'hétérogénéité de la population aussi bien au niveau des caractéristiques individuelles socio-démographiques qu'au niveau des échanges entre les individus. La transmission du VIH fait intervenir de nombreux paramètres d'ordre sociologique, démographique, épidémiologique et biologique, caractérisés par une forte variabilité. En particulier, le comportement sexuel d'un individu dépend d'un grand nombre de caractéristiques comme l'âge, le sexe, l'état de santé, la classe sociale, la période de mesure et la région considérée. L'approche mathématique prend en compte cette hétérogénéité en divisant la population en un nombre limité de sous-populations homogènes en fonction de caractéristiques principalement socio-démographiques et épidémiologiques : par exemple, le sous-groupe des homosexuels non contaminés à partenaires multiples, ou celui des hétérosexuels en couple dans lequel l'homme est infecté. Cette classification présente les inconvénients d'être arbitraire et rigide alors qu'elle est susceptible de varier dans le temps, selon les populations considérées et même selon les hypothèses choisies. L'approche mathématique est donc surtout adaptée pour étudier les tendances de l'épidémie dans de grandes populations en ne considérant qu'un nombre limité de caractéristiques.

D'autre part, l'étude de la transmission du VIH peut se faire à différents niveaux, aussi bien dans la mise au point des enquêtes sur le comportement que dans celle des modèles épidémiques. Au niveau individuel, on s'intéresse au nombre de partenaires déclarés par les individus au cours d'une période donnée. Au niveau du réseau relationnel individuel, on recherche les caractéristiques des liaisons de l'individu (liaison d'ordre 1), celles des partenaires et l'ensemble des comportements relatifs à ces liaisons. Au niveau du réseau de la population, des enquêtes spécifiques, par exemple en boule de neige, permettent de caractériser les relations d'ordre supérieur à 1 : l'ordre 2 consiste à recueillir des informations sur les réseaux individuels des partenaires de l'individu initialement choisi. La plupart des enquêtes de population n'ont permis de documenter que le

niveau 1. L'enquête ACSF permet une telle approche puisque le questionnaire recueillait certaines caractéristiques de deux partenaires avec lesquels la personne interviewée avait eu un rapport sexuel au cours des douze derniers mois. La structure du réseau relationnel est probablement un facteur important de la diffusion du VIH dans une population, comme le montre de façon schématique la figure 1 : la transmission du VIH dans le réseau des six individus est plus limitée dans le cas 1 que dans le cas 2 alors que, au niveau individuel, le nombre de liaisons par individu est le même dans les deux cas.

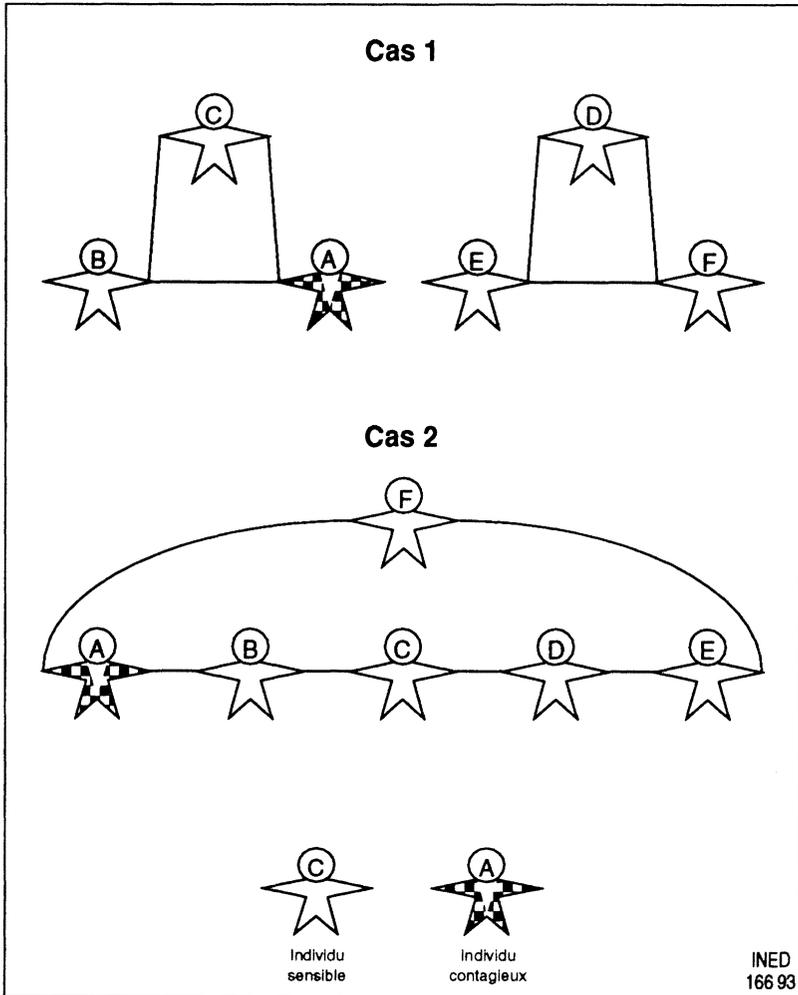


Figure 1. – Représentation de la diffusion potentielle du VIH dans un réseau relationnel

Les individus en relation sont reliés par une ligne. Les nombres de liaison sont les mêmes pour les réseaux du cas 1 et du cas 2. Cependant, dans le cas 1, au maximum trois individus peuvent être contaminés alors que, dans le cas 2, les six individus peuvent l'être

***Un modèle complet***

L'ensemble de ces remarques nous ont conduite à développer une approche de simulation individuelle qui permette de prendre en compte la diversité des comportements sexuels ou toxicomanes, le réseau relationnel et l'environnement social de l'individu, les données démographiques et la variabilité des caractéristiques biologiques. Chaque individu est représenté dans le modèle par ses caractéristiques biologiques, sociologiques et démographiques, et la vie d'un individu est simulée par une suite d'événements (rencontre d'un nouveau partenaire, modification de l'état infectieux, utilisation du préservatif selon les circonstances et le partenaire rencontré, passage au stade Sida...). Les événements sont enregistrés dans un échancier journalier qui est utilisé pour simuler la dynamique de la population : par exemple, la formation d'une nouvelle liaison sexuelle entre l'individu A et l'individu B est un événement qui induit la modification du réseau sexuel de ces individus et la prévision, dans l'échancier, de la rupture de cette liaison et de la transmission du VIH de B à A. Le fonctionnement de la simulation événementielle est présenté sur la figure 2. Lors de l'acquisition de nouvelles connaissances ou de nouvelles données concernant l'histoire naturelle de la maladie ou le comportement des individus, par exemple, de nouvelles caractéristiques individuelles ou de nouveaux événements sont intégrés au modèle. Ce type de modèle permet d'étudier l'impact du type de réseau toxicomane ou sexuel (qui a des contacts avec qui et comment s'organisent ces contacts au cours du temps ?) sur la croissance de l'épidémie, la formation et la dissolution des couples, l'existence de liaisons de couples à plusieurs partenaires, les pratiques sexuelles ainsi que les descriptions détaillées des processus biologiques (comme la variation de la probabilité de transmission au cours de l'incubation d'un individu infecté). Cette approche permet d'inclure un grand nombre de règles spécifiques du comportement sexuel et de décrire les processus de façon réaliste en prenant en compte les données numériques et déclaratives.

Un premier modèle de simulation de l'épidémie a été mis au point et a été testé à partir de données sociologiques et épidémiologiques collectées chez des homosexuels (Le Pont, 1993b). Les données provenaient en partie d'une cohorte d'homosexuels de San Francisco (évolution de l'incidence de l'infection à VIH et du nombre annuel moyen de partenaires sexuels ayant pratiqué le rapport anal réceptif avec éjaculation et sans utilisation de préservatif entre 1980 et 1988 (Hessol, 1989)). La transmission sexuelle du VIH a été simulée dans cette population définie par ses caractéristiques démographiques (taille, taux annuel de recrutement et d'émigration), épidémiologiques (prévalence initiale du VIH, distribution de la durée de séroconversion – entre la contamination et la détection possible des anticorps – d'incubation et de la durée du stade Sida, probabilités de transmission du VIH en fonction des pratiques sexuelles) et sociologiques (distribution du taux annuel de rencontre de partenaires, durées des liaisons stables et occasionnelles, pratiques sexuelles). La technique de simulation événementielle choisie a permis, en simulant le comportement individuel,

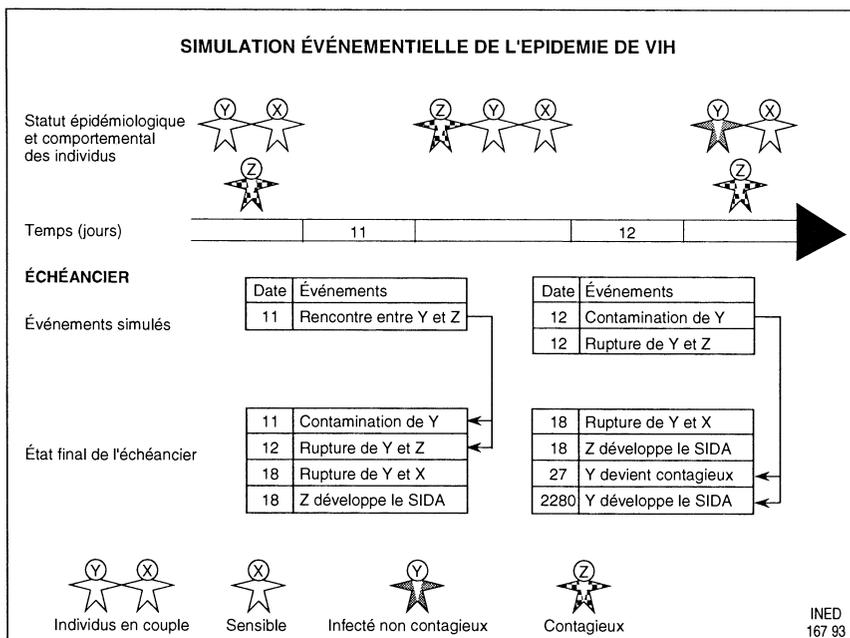


Figure 2. – Fonctionnement de l'échéancier du modèle de simulation événementiel

L'échéancier est un tableau journalier qui contient des listes d'événements classés par date croissante. Un événement est une structure qui comprend les actions à exécuter et l'objet (individu ou liaison par exemple) concerné. Chaque jour, le programme exécute les événements contenus dans l'échéancier à cette date. Ceci peut entraîner des modifications de l'état initial des individus et la création d'autres événements qui seront insérés à une date ultérieure dans l'échéancier

Au début du jour 11, les individus X et Y sensibles sont liés, Z est contagieux et célibataire. La case 11 de l'échéancier contient l'événement « rencontre de Y et Z »

Le jour 11, l'événement « rencontre de Y et Z » est exécuté. La durée de la liaison entre Y et Z est déterminée selon la loi de distribution de la durée des liaisons, et l'événement « rupture de liaison » entre Z et Y est intégré dans l'échéancier à la date adéquate (ici le jour 12, flèche descendante). Pour simuler la transmission du VIH, un chiffre est tiré aléatoirement entre 0 et 1 ; dans le cas présenté ici, ce chiffre est inférieur à la probabilité de transmission du VIH, Z transmet donc le virus à Y. La date de la contamination de Y est tirée au hasard sur la durée de la liaison entre Y et Z, et l'événement « contamination de Y » est inséré à cette date. Le jour 12, les événements prévus, « Z et Y rompent » et « Y devient infecté » sont exécutés. Les attributs des trois individus sont modifiés en conséquence et deux événements sont insérés dans l'échéancier : « Y devient contagieux » le jour 27 et « Y développe le Sida » le jour 2280. Les dates d'exécution de ces deux événements sont déterminées en fonction de la distribution du délai de latence et du délai d'incubation

d'estimer la probabilité de transmission du VIH par partenaire chez des homosexuels entre 0,04 et 0,07 (estimation proche de celle obtenue par des enquêtes épidémiologiques). L'analyse de sensibilité d'un modèle

consiste à déterminer les paramètres ayant le plus d'influence sur les résultats du modèle (c'est-à-dire les paramètres sensibles). Les résultats obtenus doivent être considérés en tenant compte du fait que l'analyse a été réalisée dans le cas d'une population particulière quant à son comportement (le taux de rencontre de partenaires par individu est passé de 20 par an en 1978 à moins de 1 en 1988, ce qui a induit une croissance très rapide de l'épidémie suivie d'une chute brutale). Il a été montré que les caractéristiques épidémiologiques suivantes ont une influence sur la croissance de l'épidémie dans cette population : la durée des stades contagieux de la période d'incubation et leur augmentation possible par le traitement de l'infection, la probabilité de transmission du VIH au cours de ces stades et la prévalence initiale du VIH. Certaines caractéristiques précises du comportement sexuel individuel sont également importantes et doivent donc être prises en compte par les modèles de simulation de la dynamique de l'infection à VIH : le nombre de partenaires par individu et par an (valeur moyenne et distribution), les modifications de comportement sexuel (nombre de partenaires et utilisation du préservatif), les dates de ces modifications et la durée des liaisons. Différents scénarios décrivant les modifications de l'activité sexuelle des individus au cours du temps ont été simulés.

Une seconde analyse a étudié l'influence de la durée des liaisons et des couples à plusieurs partenaires chez les hétérosexuels. L'objectif était de déterminer si, sous l'hypothèse d'une probabilité de transmission indépendante de la durée de la relation et du nombre de rapports sexuels (ce qui a été montré par plusieurs études épidémiologiques), la croissance de l'épidémie variait en fonction du rythme de renouvellement des liaisons et de leur ordonnancement dans le temps. Cette étude utilise donc la notion de réseau individuel. Les fréquences observées de liaisons à plusieurs partenaires ne sont pas négligeables. En effet, plusieurs enquêtes en Grande-Bretagne, en Suisse et en Suède ont montré que 16 à 20 % des individus en couple déclarent avoir eu des liaisons sexuelles en dehors du couple (Zugugin, 1989, Giesecke, 1990). La diffusion du VIH est simulée dans une population d'hommes et de femmes hétérosexuels qui ont en moyenne 2 nouveaux partenaires par an, avec une prévalence initiale du VIH de 1 %. On suppose que la probabilité de transmission est égale à 0,10, ceci pendant les périodes contagieuses situées au cours des 2-3 mois suivant la contamination et pendant les cinq années avant le développement du Sida. Les simulations sont réalisées en faisant varier les durées pendant lesquelles les individus ont un partenaire et celles pendant lesquelles ils ont deux partenaires. Les résultats montrent que la prévalence de l'infection augmente avec la durée des liaisons et, de façon plus marquée, avec la durée pendant laquelle un individu a plusieurs partenaires en parallèle. Ainsi, le pourcentage de personnes infectées par le VIH après 40 ans est multiplié par deux si chaque personne a une liaison d'un an et une d'un jour au cours de l'année, par rapport à l'hypothèse selon laquelle toutes les liaisons durent un jour (pas de relations stables).

## II. - L'épidémie en France à partir des données de l'enquête ACSF

### *Estimation rapide du taux de reproduction de l'épidémie*

En première approche, nous avons utilisé l'indicateur du risque épidémique  $R_0$  pour estimer le risque d'épidémie provenant de la transmission hétérosexuelle à partir des données ACSF (ACSF, 1992 ; Le Pont, 1993a).  $R_0$  représente le nombre moyen d'individus contaminés par un individu infecté pendant toute sa période contagieuse, dans une population où quasiment tous les individus sont indemnes du VIH (comme par exemple, au début d'une épidémie). Si  $R_0$  est supérieur à 1, chaque individu infecté transmet le virus à plus d'une personne en moyenne et le virus se propage dans la population. En revanche, si  $R_0$  est inférieur à 1, il n'y a pas d'épidémie. La définition de  $R_0$  dépend de la population considérée et du modèle utilisé (Anderson, 1988, Blower, 1990). Elle prend en compte les paramètres épidémiologiques (la probabilité de transmission au cours d'une liaison hétérosexuelle et les durées moyennes des périodes contagieuses) et le taux de rencontre de nouveaux partenaires (moyenne et variance). A partir des nombres de partenaires de sexe opposé déclarés, il a été montré, avec des faibles valeurs quant aux paramètres de transmission, qu'un taux d'utilisation du préservatif de 50 % permet au mieux de réduire mais pas d'annuler le risque d'épidémie chez les jeunes de 18 à 24 ans ( $R_0$  est toujours supérieur à 1) (Le Pont, 1993c). Ainsi, même si l'on suppose que les personnes qui ont déclaré avoir utilisé le préservatif au moins une fois au cours des douze derniers mois l'ont toujours utilisé, la valeur de l'indicateur permet de conclure que chaque personne contaminée transmet le virus à plus d'une personne en moyenne pendant toute la période d'incubation, et donc que les conditions nécessaires au démarrage de l'épidémie sont présentes.

Le modèle de simulation permettra de sélectionner les comportements qui ont une forte influence sur la diffusion de l'épidémie, et de déterminer les conditions dans lesquelles l'infection à VIH se limitera aux groupes actuellement les plus exposés (les homosexuels et les toxicomanes), ou diffusera largement dans la population générale. Les efficacités de scénarios représentant des programmes de prévention différents pourront également être comparées. Ce projet nécessite une analyse préalable des données afin de définir les entrées du modèle. De plus, des améliorations seront apportées au modèle afin de pouvoir intégrer des connaissances plus détaillées quant au comportement sexuel des individus.

### *Analyse des données ACSF*

Les données de l'enquête constituent un instantané de la structure de la population au cours de l'hiver 1991-92. On dispose ainsi de la structure démographique

graphique et d'un ensemble de réseaux individuels. Ceci permet d'initialiser la population à simuler en attribuant des caractéristiques à chaque individu (âge, sexe, lieu de naissance, vie en couple et durée de cohabitation, nombre de partenaires de chaque sexe au cours de la vie passée) et en formant les couples existants en fonction des caractéristiques des partenaires. Au début de la simulation, les individus sont tirés au sort dans l'échantillon ACSF (calé sur les résultats du dernier recensement). Pour simuler l'évolution de la population dans le temps, on dispose d'informations sur la vie sexuelle passée des individus. Ces données seront analysées pour décrire des règles de comportement qui détermineront le nombre de rencontres de partenaire de chaque sexe au cours des douze prochains mois, la durée des liaisons et les pratiques sexuelles protégées ou non, l'existence de liaisons simultanées.

On cherchera en particulier à répondre à la question « qui a des rapports sexuels avec qui ? ». Quelques données indiquant le comportement sexuel des partenaires de la personne interviewée ont été recueillies (homosexualité et multipartenariat) ; si les réponses ne permettent pas d'avoir des informations fiables sur les relations d'ordre 2, qui ne peuvent être recueillies que par des enquêtes spécifiques, elles peuvent néanmoins intervenir dans les règles de choix du partenaire. Nous utiliserons donc les informations disponibles sur les partenaires (âge, sexe, lieu de naissance, diplôme, catégorie socioprofessionnelle, niveau d'éducation, prostitution), les types de liaisons (entre conjoints, avec le partenaire principal, avec un partenaire occasionnel) et la présence de sentiments amoureux, ces informations pouvant être associées à des comportements sexuels différents. L'extraction des règles de comportement sera réalisée par des méthodes de segmentation particulièrement utiles dans le cas où l'information est abondante (Celeux, 1989).

**Amélioration du modèle** A partir de l'analyse effectuée sur les données, il est facile de compléter la fiche signalétique des individus en ajoutant par exemple la variable de la catégorie socioprofessionnelle. Un aspect méthodologique plus complexe concerne le mode de représentation des règles de comportement dans le modèle de simulation.

En effet, comment représenter de manière compréhensible par le modèle une règle qui indique combien de partenaires, de même sexe et de sexe opposé, aura dans l'année un homme de 33 ans qui a eu des rapports sexuels avec cinq femmes au cours de sa vie ? La méthode la mieux adaptée consiste à représenter ces règles sous forme de langage naturel et de les rassembler dans une base de connaissances indépendante du modèle. Un moteur d'inférence permet de sélectionner les règles à appliquer pour répondre à une question posée par le modèle au cours de la simulation (par exemple, pour déterminer la valeur d'un paramètre à utiliser ou le type d'action à exécuter). La base de connaissances et le moteur d'inférence

constituent ce que l'on appelle un système expert. Le système expert utilisé est largement inspiré du système Jonathan développé par P. Frot (Frot, 1988). Il sera en particulier important de vérifier la cohérence des règles de la base de connaissances. C'est pourquoi le système-expert peut également fonctionner non seulement en mode déductif (chaînage avant) mais aussi en mode inductif (chaînage arrière), qui consiste à rechercher si les règles qui permettent d'aboutir à une conclusion donnée ne sont pas contradictoires.

La base de connaissances doit également comprendre des règles décrivant l'évolution démographique (décès, naissance, mariage) ou biologique d'un individu (histoire naturelle de la maladie, probabilité de transmission selon les stades de l'incubation) à partir des données biologiques, épidémiologiques et démographiques publiées. Évidemment, il est nécessaire de tester si l'ensemble des règles, dont on a vérifié la cohérence, vont permettre de simuler la dynamique prévue pour la population. Ceci d'autant plus que la structure de la population est susceptible d'évoluer au cours du temps d'une façon qui peut ne pas avoir été prévue par les règles. Par exemple, la représentation du choix des partenaires doit prendre en compte l'offre et la demande de partenaires, qui peuvent évoluer au cours de la croissance de l'épidémie en raison de la mortalité par Sida (20) : ainsi, il est possible que le nombre de prostituées qui n'utilisent pas de préservatif avec leur client diminue au cours du temps. Plusieurs règles peuvent alors être proposées : les prostituées qui n'utilisent pas de préservatifs auront un plus grand nombre de clients, ou bien les clients adopteront le préservatif. Autre exemple, le nombre de partenaires des femmes âgées de 18 à 25 ans diminue alors que la demande des hommes en partenaires appartenant à cette population reste identique. Ou encore, en raison de l'évolution des comportements au cours des générations, les règles qui ont été déterminées à partir des données collectées chez les individus âgés de 45 ans au moment de l'enquête ne peuvent s'appliquer aux individus qui arrivent à cet âge au cours de la simulation (et qui avaient 35 ans au moment de l'enquête). Dans ces situations, on devra indiquer des règles de changement de comportement en fonction de la structure de la population : la création d'une base de connaissances indépendante du modèle de simulation facilite l'intégration de nouvelles informations. Par exemple, s'il s'avère que l'un des facteurs qui influe le plus sur le comportement des individus est la nature de leurs fantasmes, le modèle devrait alors prendre en compte ces fantasmes et les faire évoluer au cours de la simulation.

### *Les questions qui restent posées*

#### *1 – L'importance de la transmission de la mère à l'enfant*

Cette question nécessite d'intégrer des données sur la fécondité en fonction de l'âge et du statut sérologique, et sur le taux d'interruption de grossesse en particulier chez les femmes enceintes séropositives.

## *2 – La variabilité des comportements en fonction de la région considérée*

Elle sera prise en compte en simulant la dynamique de l'épidémie à partir de données collectées dans des régions présentant des caractéristiques différentes. Lors de l'estimation de l'indicateur du risque d'épidémie chez les hétérosexuels, on a d'ailleurs montré de fortes variations régionales (Le Pont, 1993a).

## *3 – La population toxicomane et prostituée*

En raison du faible nombre de personnes qui ont déclaré avoir une pratique toxicomane par voie intraveineuse et de femmes prostituées se déclarant telles, les données concernant les individus présentant ces caractéristiques devront être complétées par celles recueillies lors d'enquêtes plus spécifiques.

## *4 – L'épidémie vue par les hommes et vue par les femmes*

Dans l'enquête ACSF, comme dans la plupart des enquêtes sur le comportement sexuel, le nombre moyen de partenaires sexuels déclarés par les hommes hétérosexuels est supérieur au nombre de partenaires déclarés par les femmes. Il est donc possible que l'offre et la demande de partenaires ne puissent être satisfaites dans le modèle. On peut simuler l'épidémie à partir des données des hommes, des femmes ou en choisissant des valeurs intermédiaires entre les déclarations des hommes et celles des femmes.

## *5 – L'influence de la structure des réseaux relationnels*

On dispose du nombre de partenaires par individu au cours de différentes périodes et des caractéristiques des deux derniers partenaires sur les douze derniers mois. Ces informations ne permettent pas de définir un réseau avec précision. Deux études peuvent être réalisées : le modèle génère la dynamique du réseau relationnel à partir des données de l'ACSF et on étudie à intervalles de temps réguliers les caractéristiques du réseau ainsi généré. Ou bien, on fixe certaines caractéristiques du réseau, et on regarde si les épidémies obtenues diffèrent lorsque l'on fait varier ces caractéristiques. Par ailleurs, on peut simuler la structure sociale de la population, en attribuant à chaque individu des groupes relationnels représentant son entourage familial, amical, de travail, de loisir d'un individu. On fait l'hypothèse que les individus choisissent leurs partenaires dans cet entourage. Les épidémies obtenues par simulation seront comparées pour des populations de structures sociales différentes, définies par la taille, le nombre des groupes relationnels par individu et par l'évolution de la composition des groupes relationnels au cours du temps.

### 6 – Les scénarios de changement de comportement

La simulation de scénarios de changement de comportement pouvant résulter de campagnes d'éducation permettra de sélectionner ceux qui sont le plus efficaces pour réduire la diffusion du VIH. Il est également possible d'indiquer des règles de modification du comportement individuel, en fonction de la proximité de cet individu avec des personnes malades ou des personnes ayant adopté des pratiques sexuelles à moindre risque.

### 7 – L'analyse de sensibilité du modèle

Cette analyse permettra de déterminer les comportements qui ont un effet important dans l'évolution de l'épidémie. En raison de l'augmentation de la complexité du modèle, l'analyse de sensibilité du modèle nécessitera des techniques d'échantillonnage des paramètres comme celle des carrés latins. Ces techniques seront intégrées au modèle de simulation (Iman, 1981).

## Conclusion

La diffusion de l'infection à VIH détaillée selon les différents groupes d'âge et de comportements à risque sera étudiée. Il sera alors possible de déterminer les caractéristiques de la diffusion du VIH en particulier chez les jeunes, pour lesquels on a mis en évidence, en première approche, un risque d'épidémie en ne prenant en compte que la transmission hétérosexuelle (Le Pont, 1993c). Les simulations pourront aider à orienter les campagnes de prévention dans ce groupe.

La démarche qui a été mise en œuvre par l'enquête ACSF a été basée sur une collaboration importante entre chercheurs de disciplines différentes; le modèle de simulation intègre ces données, ce qui nécessite la mise au point de modèles de représentation de la transmission du VIH et de l'histoire de la maladie, ainsi qu'un modèle de représentation du comportement dont la validité devra être discutée avec des sociologues.

Françoise LE PONT

## RÉFÉRENCES

- ACSF principal investigators (1992). « AIDS and Sexual behaviour in France », *Nature*, 360 : 407-409.
- ANDERSON (R.M.), MAY (R.M.) (1988). « Epidemiological parameters of HIV transmission », *Nature*, 333 : 514-519.
- ANDERSON (R.M.), MAY (R.M.), BOILY (M.C.) *et col.* (1991). « The spread of HIV-1 in Africa : sexual contact patterns and the predicted demographic impact of AIDS », *Nature*, 352 : 581-589.

- BAILEY (N.T.J.) (1975). *The mathematical theory of infectious diseases and its applications*, London : Edition Griffin.
- BLOWER (S.M.), ANDERSON (R.M.), WALLACE (P.) (1990). «Loglinear models, sexual behavior and HIV : epidemiological implications for heterosexual transmission», *Journal of Acquired Immune Deficiency Syndromes*, 3 : 763-772.
- BLOWER (S.M.), HARTEKR (D.), DOWLATABADI (H.) *et col.* (1991). «Drugs, sex and HIV : a mathematical model for New York City», *Philosophical Transactions Royal Society of London B*, 321 : 171-187.
- CELEUX (G.), DIDAY (E.), GOVAERT (G.) *et col.* (1989). *Classification automatique des données*, Paris : Edition Dunod.
- FROT (P.) (1988). *Trois systèmes experts en Turbo C*, Paris : Edition Sybex.
- GIESECKE (J.), GÖTHBERG (M.), SCALIA-TOMBA (G.), TÜLL (P.) (1990). «Sexual contact patterns of young people in Sweden. A random survey», Résumé N° SC585, *Vth International conference on AIDS*, juin, San Francisco.
- HESSOL (N.A.), LIFSON (A.R.), O'MALLEY (P.M.) *et col.* (1989). «Prevalence, incidence, and progression of Human Immunodeficiency Virus infection in homosexual and bisexual men in hepatitis B vaccine trial; 1978-1988», *American Journal of Epidemiology*, 130 : 1167-1175.
- IMAN (R.L.), HELTON (J.C.), CAMPBELL (J.E.) (1981). «An approach to sensitivity analysis of computer models : part I-Introduction, input, variable selection and preliminary variable assessment. part II-Ranking of input variables, response surface validation, distribution effect and technique synopsis», *Journal of Quality Technology*, 13 : 174-420.
- JACQUEZ (J.A.), SIMON (C.P.), KOOPMAN (J.) (1989). «Structured mixing heterogeneous mixing by the definition of activity groups», *Lectures Notes in Biomathematics*, 83 : 301-315.
- LE PONT (F.), FLAHAULT (A.), CHAU (N.P.), GUIGUET (M.), VALLERON (A.-J.) (1989). «Modèles mathématiques de diffusion de l'épidémie de SIDA. Revue de la littérature», *Rétrovirus, La Revue du SIDA*, 2 : 107-22.
- LE PONT (F.), COSTAGLIOLA (D.), MASSARI (V.), VALLERON (A.-J.) (1990). «Size of the epidemic of transfusion-associated AIDS in France», N° FC215, *Vth International conference on AIDS*, juin, San Francisco.
- LE PONT (F.), BLOWER (S.) (1991). «The supply and demand dynamics of sexual behavior : implications for heterosexual HIV transmission», *Journal of Acquired Immune Deficiency Syndromes*, 4 : 987-999.
- LE PONT (F.) (1993a). «Prévision de l'épidémie». Chap 9. in : SPIRA (A.), BAJOS (N.), *Les Comportements Sexuels en France*, Paris : Documentation Française, 277-286.
- LE PONT (F.), VALLERON (A.-J.) (1993b). «Modélisation événementielle de la dynamique de la transmission de l'infection à VIH», *Revue scientifique et technique de l'Office international des Epizooties*, 11 : 73-82.
- LE PONT (F.), COSTAGLIOLA (D.), ROUZIUX (C.), VALLERON (A.-J.) (1993c). «Estimation du risque résiduel de transmission du VIH par transfusion sanguine en France de 1990 à 1992 et impact prévisible du dépistage systématique de l'antigène p24/25», *Revue d'Épidémiologie et de Santé Publique* (sous presse).
- LERIDON (H.) (1977). *Human fertility : the basic components*, Ed University of Chicago Press.
- PENNEC (S.) (1993). Le passage à la retraite d'une génération féminine : une projection par simulation individuelle, *Population*, 3 : 655-682.
- WHO (1992). «AIDS surveillance in Europe». *Quarterly report*, 34, 30 June, Paris : WHO Collaborating Centre on AIDS.
- ZEUGIN (P.), DUBOIS-ARBER (F.), HAUSSER (D.) *et col.* (1989). «Sexual behavior of young adults and the effects of AIDS-prevention» campaigns in Switzerland. Résumé N° WAP116, *Vth conférence sur le SIDA*, juin, Montréal.

**LE PONT (Françoise). – Vers un modèle prévisionnel de développement de l'infection à VIH en France à partir de l'enquête ACSF**

Ce chapitre présente la structure d'un modèle de prévision de l'épidémie de Sida qui prend en compte les caractéristiques socio-démographiques de chaque individu et permet de décrire sa vie par une suite d'événements (rencontre de partenaire, passage au stade Sida par exemple). Par rapport à un modèle mathématique classique qui représente des sous-populations, l'approche individuelle permet de prendre en compte la variabilité des caractéristiques du comportement sexuel qui ont été collectées par l'enquête ACSF, comme la structure initiale de la population et de son réseau sexuel, le nombre de nouveaux partenaires par an, les caractéristiques des partenaires (âge, sexe, monogamie, catégorie socio-professionnelle) et celles de la liaison (durée, en couple ou non). Ces informations peuvent être représentées dans le modèle par des règles de la forme « Si... Alors », gérées par un système expert relié au modèle. Cette approche permet de décrire des changements éventuels de comportement afin de comparer leur impact sur la croissance de l'épidémie et d'étudier la diffusion du VIH dans la population générale, pour laquelle un risque de développement d'une épidémie a été mis en évidence chez les 18-24 ans.

**LE PONT (Françoise).– Estimating a model for the spread of the HIV AIDS epidemic in France, based on the ACSF survey**

In this chapter an estimated model of the AIDS epidemic is presented. It takes into account the socio-demographic characteristics of each individual, whose life-history is described by a series of events, e.g. encountering a partner, or being diagnosed as suffering from AIDS. By relating this to a classic mathematical model for sub-populations, the individual approach makes it possible to take account of the variability of sexual behaviour, documented in the ACSF survey, as well as the initial structure of the sexual network that the individual belongs to (the number of new partners per year, the characteristics of the partners such as age, sex, monogamy, socio-occupational group, and of the affairs (duration, cohabitation, etc.)). These data can be presented in the form of model using an 'if...then' approach which makes it possible to describe possible changes in behaviour and to compare their impact on the growth of the epidemic, and to study the spread of HIV in the general population, in which the risk of an epidemic that would effect the 18-24 age group has been clearly identified.

**LE PONT (Françoise). – Hacia un modelo de previsión de la epidemia en Francia a partir de la encuesta ACSF**

Este capítulo presenta la estructura de un modelo de previsión de la epidemia del Sida. Este modelo tiene en cuenta las características socio-demográficas de cada individuo y permite una descripción de su vida a través de una serie de acontecimientos (encuentro de pareja, paso a la etapa del sida, por ejemplo). En comparación con un modelo matemático clásico que representa a sub-poblaciones, la perspectiva individual permite la toma en consideración de gran variedad de características del comportamiento sexual; estas características, que incluyen la estructura inicial de la población y de su grupo de relaciones sexuales, el número de nuevas parejas por año, las características de tales parejas (edad, sexo, monogamia, categoría socio-profesional) y las de la relación (duración, en pareja o no), son las recogidas por la encuesta ACSF. Estas informaciones se pueden representar en el modelo a través de reglas de la forma "Si...Entonces" organizadas según un sistema experto ligado a él. Esta perspectiva permite la descripción de cambios eventuales de comportamiento con el fin de comparar su impacto sobre el crecimiento de la epidemia y de estudiar la difusión del virus en la población general. Un ejemplo de riesgo de desarrollo de una epidemia se ha llevado a cabo para el grupo de 18-24 años.