

simulation-modélisation

L'ÉPIDÉMIE DE SIDA

UNE MODÉLISATION

Anne Florimond

PROFESSEUR DE SVT
LYCÉE RICHELIEU, RUEIL-MALMAISON

Nathalie Noris

PROFESSEUR DE SVT
LYCÉE RICHELIEU, RUEIL-MALMAISON

José Quaresma-Nunes

PROFESSEUR DE SVT
LYCÉE MARIE-CURIE, SCEAUX ERTÉ ACCESS IRNRP

Faire des prédictions sur l'évolution future de l'épidémie de sida est un enjeu majeur. Ces prédictions doivent s'appuyer sur des modèles numériques à compartiments qui mettent en équation des flux d'individus et qui doivent pouvoir être testés avec des données accessibles.

Le logiciel Vensim¹ permet une modélisation compartimentale déjà largement exploitée dans le domaine des sciences de la Terre². Pour autant, la biologie, et particulièrement l'immunologie, peut à son tour se prêter à la construction de modèles à compartiments. En terminale S, la confection avec le logiciel Vensim de modèles plus ou moins complexes a pour intérêt, au-delà d'une discussion sur la valeur prédictive des modèles, de sensibiliser à l'impact des conduites et des thérapies sur l'évolution future de l'épidémie de sida.

La naissance du modèle : une construction intuitive et mathématique

Classiquement proposée aux élèves, la courbe donnant les différents états qui se succèdent chez une personne infectée va ici nous permettre de définir le nombre de compartiments qui figureront dans notre modèle à cette échelle. On perçoit aisément qu'à l'échelle d'une population il y a quatre compartiments possibles, qui renferment respectivement des individus sains, des individus en phase de primo-infection, des individus en phase asymptomatique et des individus atteints de sida. La courte durée de la primo-infection par rapport à celle de la phase asymptomatique nous autorise à réunir dans un même compartiment les personnes qui sont dans l'un de ces deux états. Notre modèle comportera donc trois compartiments (figure 1) :

- U (« Uninfected ») : nombre d'individus non infectés dans la population ;
- L (« Latent ») : nombre d'individus infectés mais n'exprimant pas encore le sida ;

– S (« Symptomatic ») : nombre d'individus atteints de sida (ce nombre inclut les personnes décédées du sida).

Nous devons ensuite représenter les flux entre les compartiments, car notre modèle devra calculer les variations d'effectifs au cours du temps dans chaque compartiment. Le pas de temps sera le mois. On peut facilement concevoir deux flux (figure 2) :

– un flux du compartiment L vers le compartiment S. Comme la phase asymptomatique dure en moyenne 8 ans, on peut établir que le « taux de passage » de L vers S vaut $1/(8 \times 12)$ soit 0,010417 « par mois ». Chaque mois, le compartiment S s'enrichit d'un nombre de personnes appelé dS/dt , qui vaut 0,010417 L. À l'inverse, le compartiment L s'appauvrit du même nombre de personnes ;

– un flux du compartiment U vers le compartiment L, dont la quantification paraît problématique. Celle-ci nécessite de faire appel à des équations mathématiques plus complexes³. La variation du nombre de personnes par mois s'exprime de la manière suivante : $dU/dt = -p_1(U/L)$, p_1 étant la probabilité pour qu'un individu sain devienne infecté ; p_1 dépend du nombre de rapports sexuels par partenaire, du nombre de partenaires par individu et de la probabilité pour qu'un rapport sexuel soit contaminant.

Nous en sommes là au stade d'un modèle « sur papier » qui n'est pas encore un modèle calculatoire, car il nous manque des données pour paramétrer le modèle. On a besoin de connaître d'une part le paramètre p_1 pour le calcul du taux de passage de U vers L à chaque pas de temps (chaque mois), d'autre part les effectifs initiaux dans chaque compartiment afin de pouvoir initier le calcul, de mois en mois, de l'évolution des effectifs dans chaque compartiment.

À ce stade, nous ne pouvons donc pas aller plus loin dans la confection du modèle.

Le fonctionnement du modèle

Un modèle à calibrer

Pour calibrer le modèle et le faire fonctionner, nous utilisons une population particulière et qui a été bien suivie au niveau statistique. Il s'agit de sujets masculins, tous homosexuels, dans l'État du Massachusetts. Pour

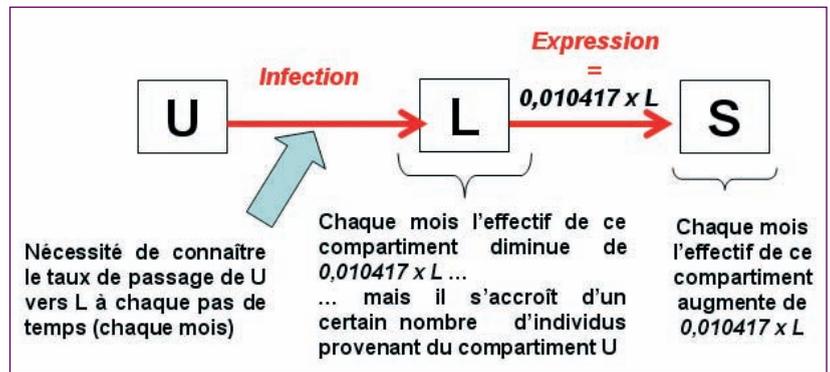
1. Vensim PLE est gratuit et téléchargeable à l'adresse : www.vensim.com

2. Les *Dossiers de l'ingénierie éducative*, n° 53, décembre 2005, (« Réchauffement de la planète : l'Homme est-il responsable » et « Vensim, un outil généraliste de modélisation dynamique »), ainsi que les nombreuses utilisations pédagogiques du logiciel Vensim proposées par l'équipe ERTÉ ACCES d'Orléans-Tours : <http://aces.inrp.fr/aces/equipes/modelisation/travaux/CCCIC/>

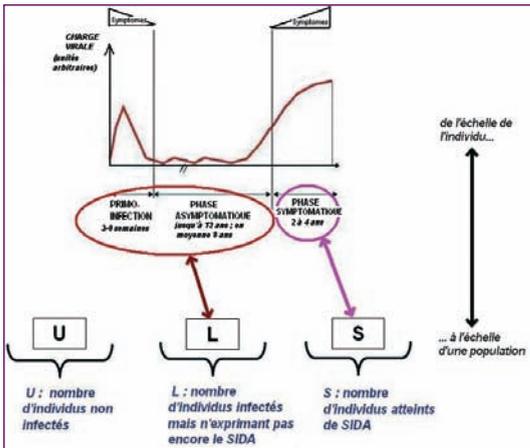
3. Les détails de la mise en équation des flux sont communiqués dans un article « référent » : « Simplicity vs complexity in deterministic models : an application to AIDS data » (S. SANDBERG, T. AWERBUCH et R. GONIN, 1994).

4. Cette fonctionnalité nécessite la version payante Vensim PLEplus, qui offre la possibilité d'ouvrir directement des banques de données au format xls.

POSSIBLE?



2. Une ébauche de modèle « sur papier » pouvant être construit par les élèves.



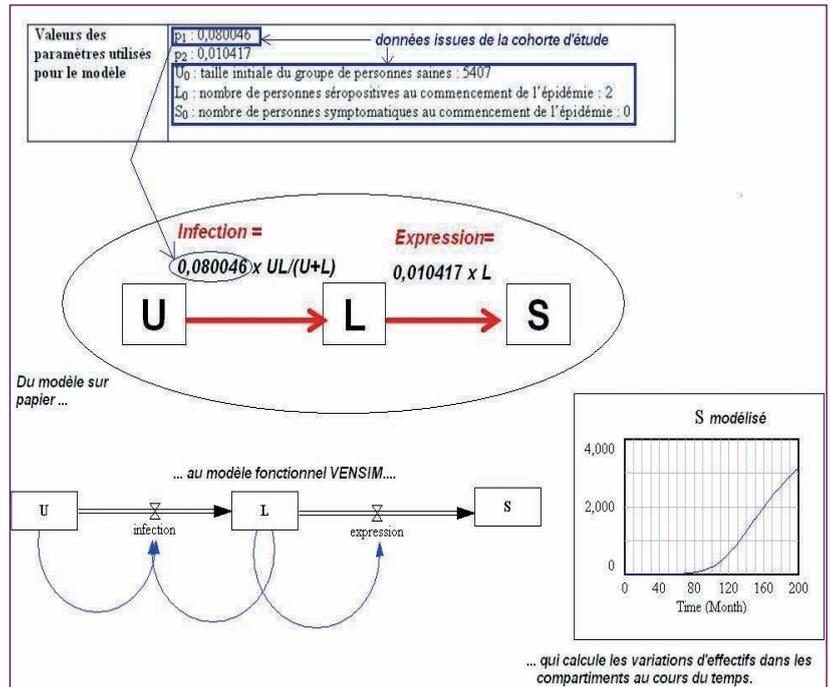
1. Passage de l'échelle de l'individu à l'échelle populationnelle : naissance du modèle à compartiments.

cette population, on dispose d'un jeu de données ou d'extrapolations (effectifs initiaux dans les trois compartiments, probabilité de transmission de l'infection). Cette population sera notre cohorte d'étude.

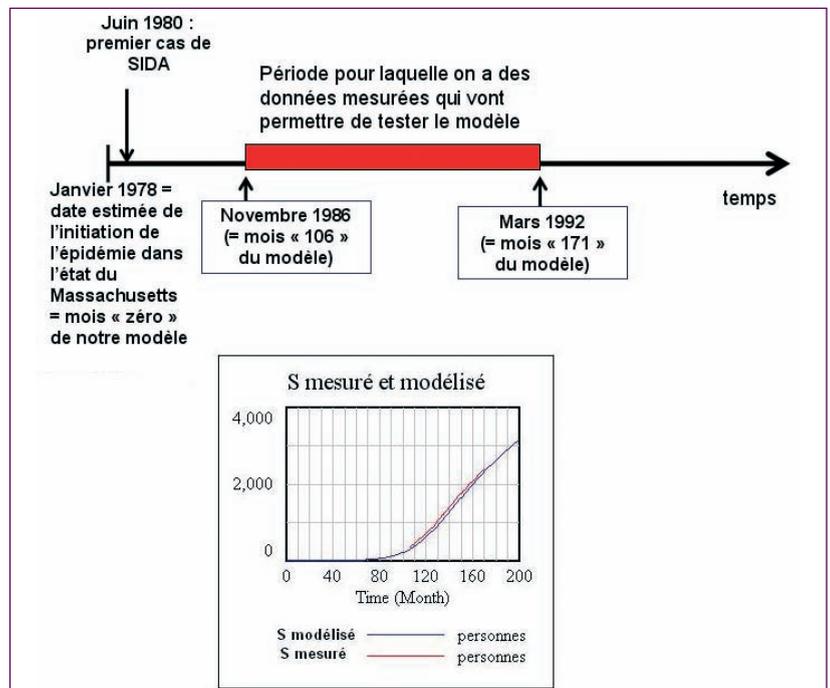
En entrant les valeurs de p_1 et les effectifs initiaux, la mise en équations des flux entre les compartiments peut désormais être finalisée, et le modèle peut être initialisé. À ce stade, nous détenons un modèle qui fonctionne (figure 3), puisqu'il calcule les variations au cours du temps des effectifs dans les différents compartiments. Afin de tester sa validité, il est nécessaire de comparer les effectifs calculés par le modèle aux effectifs réels de la cohorte d'étude.

Un modèle à tester par confrontation au réel

Le modèle doit être testé avec un jeu de données issu de la même cohorte d'étude que celle ayant servi à son paramétrage. Nous exploitons donc un extrait du registre de statistiques de l'État du Massachusetts (données obtenues auprès du Département de Santé publique du Massachusetts), qui fournit l'évolution, mois par mois, des effectifs de personnes pour lesquelles le sida a été diagnostiqué entre novembre 1986 et mars 1992. Nous prenons comme échelle de temps le nombre de mois écoulés depuis la date estimée (janvier 1978) de l'initiation de l'épidémie dans cet État. En intégrant ces données (préalablement formatées en feuille de calcul⁴) à notre modèle précédent, on peut estimer la pertinence du modèle (figure 4).

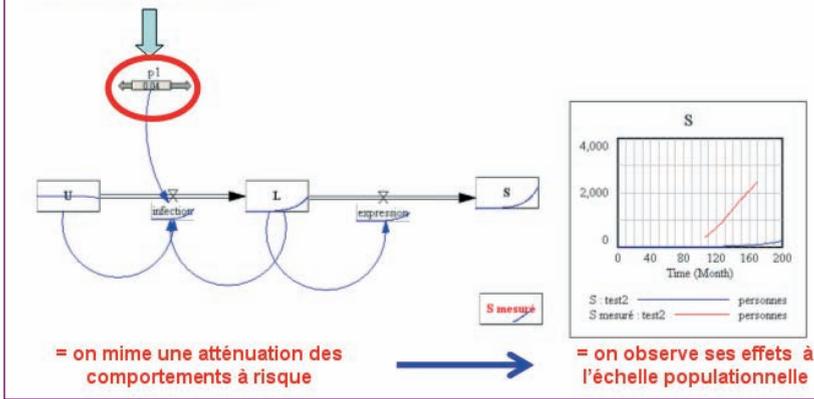


3. Un premier modèle calculatoire élaboré avec le logiciel Vensim.



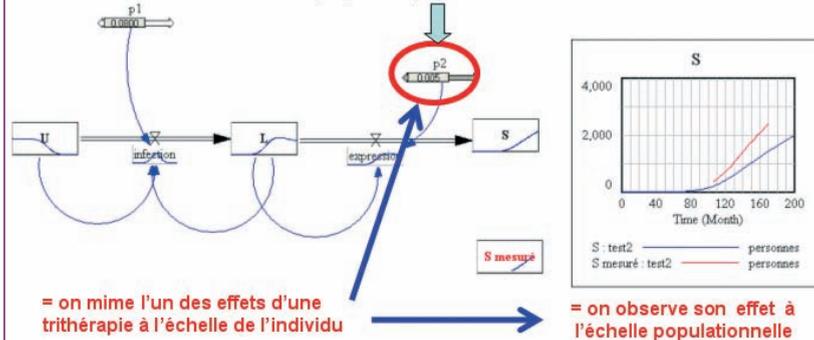
4. Mise à l'épreuve du modèle : place du jeu de données (haut de la figure) et confrontation de « S modélisé » et « S mesuré » (bas de la figure).

On diminue la valeur de p_1 (ici 0,04 au lieu de 0,080046)



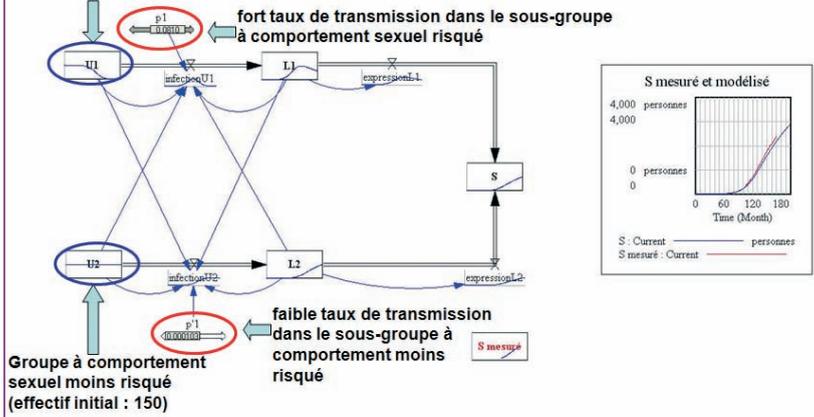
5. Un modèle où le paramètre « probabilité de transmission de l'infection » est modulable.

on règle p_2 à 0,005 ce qui correspond à une phase asymptomatique de 16 ans

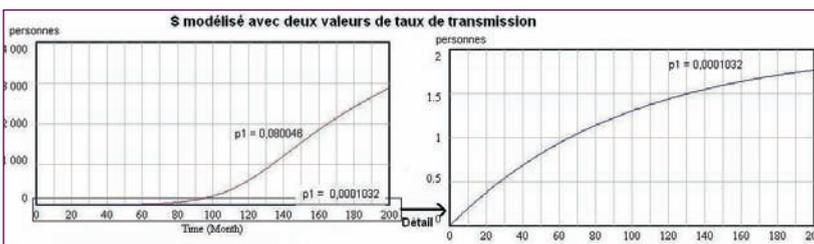


6. Un modèle où le paramètre « taux de passage d'un individu du compartiment infecté au compartiment symptomatique » est modulable.

Groupe à comportement sexuel risqué (effectif initial : 4872)



7. Un modèle plus compartimenté introduisant une hétérogénéité comportementale.



Ici, nous observons une bonne adéquation entre l'évolution temporelle du nombre de sujets atteints de sida, calculée par le modèle, et la réalité. Le modèle explique donc *a posteriori* le scénario de l'évolution de la maladie dans la population étudiée. Il donne accès aux paramètres qui caractérisent la maladie.

La complexification des modèles au service de l'éducation à la santé

Modifier le paramétrage du modèle pour sensibiliser à l'impact des conduites et des thérapies

Dans notre modèle référent, le paramètre p_1 (probabilité pour qu'un individu sain devienne infecté) avait été fixé à 0,080046. Il s'agissait de la probabilité estimée de transmission de l'infection dans notre cohorte d'étude. En intégrant à notre modèle un curseur offrant la possibilité de faire varier p_1 , tout se passe comme si on simulait des changements de comportement. Notre cohorte d'étude étant une population à risque, il paraît logique de diminuer la valeur de p_1 afin de mimer une atténuation des comportements à risque. Il est alors possible d'en apprécier les effets (figure 5).

Toujours dans notre modèle référent, le paramètre p_2 (taux de passage d'un individu du compartiment infecté au compartiment symptomatique) avait été fixé à 0,010417. Cela correspondait à une phase asymptomatique de 8 ans. En intégrant au modèle un nouveau curseur offrant la possibilité de faire varier p_2 , tout se passe comme si on simulait une modification de la durée de la phase asymptomatique. Ici, diminuer p_2 revient à simuler un allongement de la phase asymptomatique. Concrètement, cela mime l'action d'une trithérapie. Il est alors possible d'en quantifier les effets (figure 6).

Sous-compartimenter le modèle pour simuler une hétérogénéité comportementale

Le modèle de départ est basé sur 3 compartiments U (sains), L (infectés mais non malades) et S (symptomatiques). Il considère, par principe, puisque c'est un modèle à compartiments, que chaque compartiment est formé d'unités homogènes (les individus) qui ont tous un même comportement, régi par les équations et paramètres mathématiques. Or on sait bien que les individus ont des comportements différents les uns des autres vis-à-vis du risque même s'ils font partie du groupe à risque. Il n'est pas possible de se placer au niveau de l'individu avec un modèle à compartiments, mais on peut créer des sous-groupes :

- un premier sous-groupe d'individus à comportement risqué : il est désigné par « U1 » dans le modèle informatique et aura une probabilité p_1 égale à 0,081088, supérieure à la valeur de p_1 du modèle sans sous-groupe ;

— un second sous-groupe d'individus à comportement moins risqué : désigné par « U2 » dans le modèle informatique, il aura une probabilité $p'1$ beaucoup plus faible, égale à 0,0001032.

Un tel modèle informatique rend compte des variations réellement observées du nombre de personnes atteintes du sida (figure 7).

Utiliser le modèle à sous-compartiments pour sensibiliser les populations à risque

Dire que certains individus ont un comportement sexuel plus risqué que d'autres revient à dire qu'ils ont un plus grand nombre de partenaires sexuels ou un plus grand nombre de rapports sexuels ou des pratiques ayant une plus forte probabilité de transmission du virus que les autres individus. On peut espérer agir sur ce groupe par les messages de prévention pour faire diminuer la probabilité de transmission du virus en augmentant l'usage des préservatifs. Si la politique de prévention atteint son but, on peut alors arriver à une situation où le groupe à comportement risqué U1 rejoint le groupe à comportement sexuel moins risqué U2. Dans ce cas, on fait chuter le nombre de cas de sida de manière spectaculaire et quantifiable (figure 8).

Conclusion

Dans le domaine de la modélisation en épidémiologie, les modèles Vensim qui viennent d'être décrits sont des modèles déterministes : ce sont les conditions initiales qui déterminent l'état futur du système et permettent ainsi de modéliser l'état de l'évolution des populations. Il existe d'autre part des modèles stochastiques, qui, eux, sont probabilistes et modélisent l'évolution de phénomènes à l'échelle de l'individu. Ce sont alors d'autres outils que Vensim qui permettent la construction de ce type de modèles. Ainsi, la modélisation et ses divers aspects (modèles déterministes ou modèles stochastiques, modèles à compartiments ou modèles bien centrés sur l'individu) trouvent bel et bien leur place en immunologie. ●

Si la politique de prévention atteint son but, on peut alors arriver à une situation où le groupe à comportement risqué rejoint le groupe à comportement sexuel moins risqué. Dans ce cas, on fait chuter le nombre de cas de sida de manière spectaculaire et quantifiable.

Bibliographie

- « Simplicity vs complexity in deterministic models: an application to AIDS data » (S. SANDBERG, T. AWERBUCH et R. GONIN, 1994).
- « Une épidémie mise en équations », *La Recherche* n° 335, octobre 2000, J.-P. Gouteux, M. Artzrouni, M. Jarry.

La « quête de données »

Au-delà de la connaissance de l'évolution d'une pathologie ou de son incidence à un instant t dans une population, la recherche de données est essentielle dans la construction d'un modèle. Les données récoltées peuvent aider à son élaboration en guidant le modélisateur dans ses choix mais surtout elles permettent une validation du modèle en le confrontant au réel. Il s'agit là d'une étape importante de mise à l'épreuve qui permet de dépasser le modèle purement « théorique ».

Les données recherchées peuvent être issues d'articles de chercheurs, comme cela a été le cas dans l'exemple traité ici à partir de la publication de S. Sandberg, T. Awerbuch et R. Gonin dans *J. Biol. Systems*. Il est cependant difficile de trouver des données suffisamment variées et complètes utilisables dans le cadre de la modélisation choisie. L'accès à ces données n'est pas toujours aisé et nécessite souvent d'avoir un contact auprès de chercheurs qui pourront orienter cette recherche. Le développement sur la Toile de nombreuses banques¹ de données offre de nouvelles perspectives.

Quelques sites de référence proposent des données démographiques générales sur les populations, la natalité, la mortalité avec des informations classées par sexe et par âge. C'est le cas de la Division de la Population des Nations unies² au niveau mondial ou encore de Human Mortality Database³, base mise en place en 2000 par l'université de Californie à Berkeley (USA) et l'Institut Max-Planck pour la recherche démographique à Rostock (Allemagne). En France, l'INSEE⁴ et l'INED⁵ fournissent des éléments comparables.

Même si ces deux derniers sites apportent également des données concernant les causes de décès, la consommation de drogues, de tabac, d'alcool ou sur les consultations médicales, d'autres, plus spécialisés, permettent d'obtenir des données épidémiologiques plus précises.

On retiendra parmi eux :

- la banque de données de l'Organisation mondiale de la santé⁶ qui apporte des statistiques pour les 193 États membres ;

- les bases de données accessibles depuis le site du Bureau régional de l'Europe de l'OMS comme : HFA-DB⁷, la base de données européennes de la santé pour tous, renfermant des données sur 600 indicateurs pour les 53 États membres. Sa mise en place étant ancienne, elle présente l'avantage de contenir des données remontant à 1970 pour certains indicateurs ce qui permet une étude sur une longue période ; HFA-MDB⁸, qui présente des indicateurs de mortalité répartis en 67 causes de décès, par âge et sexe ;

- des bases de données spécialisées dans le domaine du sida : EuroHIV⁹, qui coordonne la surveillance du VIH/sida depuis 1984 en Europe et dont le site renvoie vers deux sources de données importantes : la base du système d'information sur les maladies infectieuses (CISID) et la base européenne de prévalence du VIH ; UNAIDS¹⁰, site des Nations unies dédié au sida connu en français sous le nom Onusida dont la base renvoie à celle de l'OMS mais avec uniquement les données concernant le sida.

Il existe bien d'autres bases de données encore, notamment françaises (Score Santé, CépIdc...), il n'est pas question ici d'en faire une liste exhaustive.

Toutes fournissent des données consultables en ligne, généralement sous forme de tableaux mais parfois aussi de graphes ou de cartes. Elles permettent également une exploitation hors ligne par l'intermédiaire soit de fichiers compressés renfermant la totalité de la base, soit de fichiers exportables dans des formats compatibles avec des tableurs, ce qui permet une utilisation ultérieure éventuelle dans des logiciels de modélisation comme Vensim. La limite d'utilisation actuelle de ces bases est liée à leur relative nouveauté. Les données proposées sont souvent récentes, notamment dans le cas de l'infection par le VIH, ce qui limite leur mise en perspective. De plus les informations sont encore trop souvent ponctuelles pour bon nombre de pays, chacun n'ayant pas suivi les mêmes procédures ou réglementations. Les bases en ligne sont un outil intéressant mais qui reste tributaire du recensement des données sur le terrain.

Toutefois l'offre d'utilisation en ligne s'est considérablement développée ces dernières années. Dans certains cas il est même possible de faire des requêtes personnalisées et de générer des cartes et des tableaux. Dans le même temps le développement d'outils parfois proposés à côté de la base de données (comme le logiciel DPS sur la base HFA-DB) traduit l'intérêt grandissant pour ce type d'informations et peut ouvrir la voie à de nouvelles exploitations pédagogiques.

1. <http://acces.inrp.fr/acces/ressources/sante/epidemiologie/donnees-epidemiologiques-sida/instrumenter/bases-de-donnees-epidemiologiques-dans-le-monde>

2. <http://esa.un.org/unpp/>

3. <http://www.mortality.org/>

4. http://www.insee.fr/fr/ffc/chifcle_fiche.asp?ref_id=NATFPS06205

5. <http://www.ined.fr/fr/>

6. <http://www.who.int/whosis/fr/index.html>

7. <http://www.euro.who.int/hfad>

8. http://www.euro.who.int/InformationSources/Data/20011017_1

9. http://www.eurohiv.org/index_fr.htm

10. <http://www.unaids.org/en/KnowledgeCentre/HIVData/Epidemiology/epidatabases.asp>