

1/ OBJECTIFS ET PROBLEMATIQUES DE LA BASE DE DONNEES HOSPITALIERE FRANCAISE SUR L'INFECTION À VIH AU MOMENT DE LA CREATION ET EVOLUTION(S) DEPUIS LA CREATION

En réponse à l'émergence de l'épidémie de VIH/SIDA, le Ministère de la Santé a créé les CISIH, Centre d'Information et de Soins de l'Immunodéficience Humaine, c'est-à-dire des centres de références localisés dans les CHU et ayant pour mission d'assurer les soins en associant les unités cliniques, médico-techniques et de recherche, de développer la recherche épidémiologique et médico-économique et d'animer les réseaux de soins. Dès ce moment des techniciens d'études cliniques ont été mis à disposition des CISIH. Parallèlement, 2 systèmes d'information spécifiques ont été mis en place. Il s'agissait d'une part du DMI, système centré sur le recours hospitalier, à visée médico-économique, basé sur le PMSI et mis à disposition des CISIH par la DHOS en 1988. L'autre système, le DMAC était basé sur le patient et ses visites à l'hôpital, à visée d'épidémiologie clinique et mis à disposition des CISIH en 1989 par le SC4 de l'INSERM, créé pour cette mission. Le Pr AJ Valleron avait été missionné pour développer ce projet par le conseiller SIDA du gouvernement, Alain Pompidou. En 1992 les deux systèmes ont été unifiés, autour du logiciel DMI2, centré sur le patient, mis à disposition des CISIH conjointement par la DHOS et l'INSERM SC4.

Les données collectées dans le DMI2 ont permis de créer en 1992 la base de données hospitalière française sur l'infection à VIH. La base a pour but de recueillir les informations cliniques sur les patients infectés par le VIH et suivis dans des Centres d'Informations et de Soins sur l'Immunodéficience humaine (CISIH). Elle est maintenant gérée au sein de l'unité INSERM UMR S 720, " Epidémiologie clinique et thérapeutique de l'infection à VIH ", qui a fait suite au SC4.

Un des atouts majeurs de la base est sa taille. Ainsi, alors même que les incidences de toutes les manifestations opportunistes ont beaucoup diminué, nous pouvons continuer à étudier l'histoire clinique de la maladie et mettre en relation les résultats biologiques et cliniques. Nous pouvons aussi étudier la survenue de pathologies encore rares, mais dont on craint qu'elles ne soient liées à l'exposition aux traitements.

À côté des travaux de recherches, la base permet de décrire l'épidémiologie de l'infection à VIH prise en charge à l'hôpital en France comme l'illustre le chapitre épidémiologie du rapport Yéni 2006¹ et contribue à l'évaluation de l'intérêt de Santé publique des antirétroviraux. Compte tenu de sa couverture (de l'ordre de 50 % à 60 % des patients pris

¹ Yéni P. Prise en charge médicale des personnes infectées par le VIH. Rapport 2006. Recommandations du groupe d'expert. In. Paris: Médecine-Sciences Flammarion; 2006.

en charge)² et de sa représentativité (caractéristiques des patients très semblables à celles observées dans VESPA qui a inclus dans des centres participant à la base comme dans des centres n'y participant pas)³, la base est une source précieuse pour décrire qui sont les patients infectés par le VIH et pris en charge à l'hôpital et comment leurs caractéristiques évoluent au fil du temps. De plus comme les antirétroviraux (62 %) et les hospitalisations (31 %) représentent 93 % des coûts de prise en charge du VIH⁴, les données disponibles permettent de caractériser l'essentiel des coûts liés à la prise en charge des patients.

Les objectifs sont d'étudier l'histoire naturelle et sous traitement de l'infection par le VIH et se déclinent en :

- contribuer à la surveillance épidémiologique de l'infection à VIH en France
- étudier l'histoire clinique de la maladie et mettre en relation les résultats biologiques et cliniques
- étudier la survenue de pathologies rares, susceptibles d'être liées à l'exposition aux traitements ou à l'infection à VIH

La Base comprend les données de 62 hôpitaux répartis dans 29 des 30 Centres d'Information et de Soins de l'Immunodéficience Humaine (CISIH). Les données, relatives à 114 548 personnes sont incluses dans la base, dont 107 742 personnes suivies au moins une fois entre le 1^{er} janvier 1992 et le 30 juin 2006, en moyenne pendant 60 mois (suivi médian 55 mois).

2/ FONCTIONNEMENT ET ORGANISATION DE LA BASE

- **Recrutement des patients**

Il s'agit d'une cohorte hospitalière, ouverte, multicentrique, ayant inclus des patients à partir de 1989. Pour être inclus dans la base, les sujets doivent répondre à trois critères : être infectés par le VIH-1 ou le VIH-2, être suivis dans un centre participant (62 hôpitaux à l'heure actuelle), et avoir donné leur consentement éclairé par écrit.

² Bernillon P, Lievre L, Pillonel J, Laporte A, Costagliola D. Record-linkage between two anonymous databases for a capture-recapture estimation of underreporting of AIDS cases: France 1990-1993. The Clinical Epidemiology Group from Centres d'Information et de Soins de l'Immunodéficience Humaine. *Int J Epidemiol* 2000;29(1):168-74.

77 449 patients relevant de l'ALD 30 VIH au 30 octobre 2004 d'après les données de la CNAM à comparer aux 48 333 patients suivis en 2005.

³ Lert F, Obadia Y, Equipe de l'étude Vespa. Comment vit-on en France avec le VIH/sida? *Population et sociétés* 2004;406:1-4.

⁴ Points de repère n°3 - Octobre 2006 - Coût de trente affections de longue durée pour l'Assurance Maladie

- Suivi des patients

Les informations sont recueillies à chaque hospitalisation (classique ou de jour) et à chaque consultation du sujet, si un événement clinique et/ou thérapeutique s'est produit ou au moins tous les 6 mois.

Si un patient change de centre mais continue à être suivi dans un centre participant à la base, le numéro d'anonymat (voir ci-dessous) permet de le repérer dans son nouveau centre. Une étude conduite sur l'année 1999, a montré un taux de perdus de vue pour les patients suivis cette année-là de 8,5% d'un niveau comparable à celui observé dans les autres cohortes françaises (APROCO-COPILOTE, AQUITAINE) ou internationales (EUROSIDA, ART-CC) mais plus faible que dans plusieurs cohortes américaines (par exemple 16,5 % dans la Base de données de AmFAR). Un article paru dans *J Clin Epidemiol*⁵, étudie le taux de perdus de vue et les facteurs associés au fait d'être perdu de vue au sein de la base. Sur 34835 patients suivis en 1999, 1478 sont décédés et 2950 perdus de vue (16,8% % parmi ceux dont le diagnostic de l'infection à VIH datait de moins d'un an et 7% parmi ceux dont l'infection à VIH datait de plus de un an). Parmi les patients de diagnostic récent, 3 facteurs sont associés à un plus grand risque d'être perdus de vue, appartenir à un autre groupe de transmission que le groupe homosexuel/bisexuel, ne pas avoir eu une pathologie classant SIDA et être un migrant (soit originaire d'Afrique subsaharienne, soit d'un pays de l'Union Européenne). Parmi les patients de diagnostic plus ancien, les patients suivis en Guyane ou en Guadeloupe sont plus souvent perdus de vue ainsi que les migrants d'Afrique sub-saharienne et les patients du groupe de transmission usagers de drogue par voie intraveineuse. La proportion de perdus de vue diminue avec l'âge et est plus élevée chez les patients ayant un faible taux de CD4 et une charge virale élevée.

- Recueil des données

Les données sont collectées à partir des dossiers médicaux. Elles incluent des données invariables (groupe de transmission, date de première sérologie positive, notion de contage daté, antécédents cliniques et thérapeutiques de l'infection à VIH, ...), des données cliniques (pathologies classantes, autres diagnostics), biologiques (charge virale et CD4, ...), et thérapeutiques (antirétroviraux (ARV), prophylaxies primaires et secondaires, traitements des manifestations cliniques). Depuis 2005, sont également collectés les consommations

⁵ Lanoy E, Mary-Krause M, Tattevin P, Dray-Spira R, Duvivier C, Fischer P, Obadia Y, Lert F, Costagliola D and the Clinical Epidemiology Group of the French Hospital Database on HIV infection. Predictors identified for losses to follow-up among HIV-seropositive patients. *J Clin Epidemiol* 2006; 59(8): 829-835.

d'alcool et de tabac, un module VHC pour les patients coinfectedés, le bilan lipidique, les transaminases et le ou les motifs d'arrêt des ARV.

En dehors de cas particulier, les données sont recueillies à partir des dossiers médicaux puis saisies via le logiciel DMI2, propriété du Ministère en charge de la Santé, par les Techniciens d'Études Cliniques (TEC) en poste dans les différents hôpitaux. Périodiquement (deux fois par an), les TEC extraient les données à destination de l'U720. Lors de cette extraction, les sujets sont indexés par un numéro d'anonymat construit à partir du nom, du prénom et du jour et du mois de naissance. Les informations sont cryptées, stockées sur des disquettes et transmises à l'U720. Les clés de cryptage (permettant le décryptage des données) sont directement communiquées à l'U720.

Avant d'être incorporées dans la base, les données sont soumises à des procédures de validation permettant d'éliminer les informations inexploitablees (élimination des doublons intra-centres, vérification des variables dont la cohérence est jugée indispensable à l'inclusion du sujet dans la base ou de toute fiche de suivi le concernant). Les doublons inter-centres (personnes ayant le même numéro d'anonymat dans des centres différents) sont également identifiés. On s'appuie alors sur la cohérence des données pour déterminer si les numéros d'anonymat identiques correspondent à un seul et même patient (vrais doublons) ou s'il s'agit de personnes différentes (faux doublons). Les données, intégrées dans une base gérée sous Oracle® version 8.0, sont analysées à l'aide du logiciel SAS® version 9.1.

L'U720 organise régulièrement un audit des données contenues dans les bases locales du DMI2, actuellement sous la responsabilité de Souhila Saidi. Il s'agit de contrôler la qualité des données incluses localement dans les CISIH. Cet audit se fait sous forme d'une ou deux journées de visite sur site par un Médecin d'Études Cliniques. Cet audit porte sur une quinzaine de dossiers tirés au sort (données invariables, pathologies classantes pour le SIDA, traitements antirétroviraux et prophylactiques, taux de CD4+ et charge virale). Plus récemment l'audit de plusieurs diagnostics non classants SIDA a été rajouté (infarctus par exemple). La conformité ou non de chacune de ces données par rapport aux dossiers patients est notée sur une fiche de recueil standardisée et un rapport d'audit est adressé à chaque centre.

- Aspects réglementaires et éthiques :

Le projet a reçu un avis favorable de la CNIL le 27 novembre 1991 (J.O. du 17 janvier 1992). En, raison de la nécessité de retourner au dossier du malade pour pouvoir valider des données dans le cadre de projets de recherche précis, un projet a été déposé à la CNIL, autorisant la mise en œuvre d'une liste de correspondance locale entre numéro d'anonymat

du patient et son identité, permettant ce retour au dossier. L'accord de la CNIL a été obtenu en 1999.

- Modalités de transmission des résultats aux patients

Une réunion annuelle est organisée avec le TRT5, qui regroupe les principales associations de malades, pour faire un bilan des travaux en cours et discuter des thèmes à développer. Il a été décidé de proposer au TRT5 de nommer un représentant au sein du CS à partir de mars 2007.