

## BE Statistique & Santé

### Infection par le VIH : Causes d'arrêts de traitement

Ce projet est à effectuer par groupe de 2 étudiants.

Vous devrez me rendre, pour le vendredi 10 février 2017, un rapport (de 30 pages maximum, hors annexes). Vous pouvez effectuer ce projet avec les logiciels de votre choix (SAS, R, ...).

## 1 Contexte du projet

Chez un patient infecté par le VIH, plusieurs phases de la maladie se succèdent :

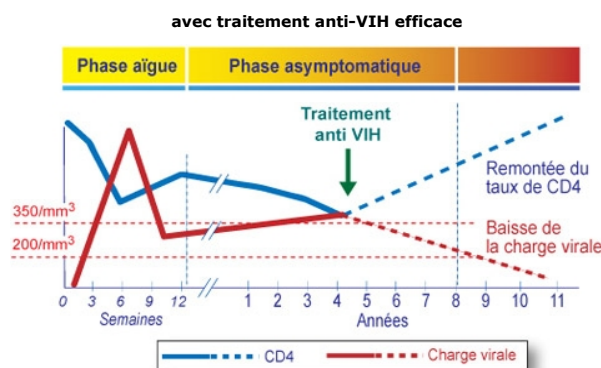
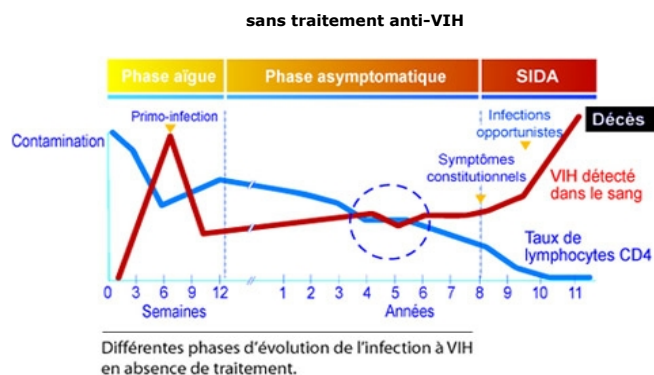
- Une première phase est appelée **phase de séroconversion** au cours de laquelle les anticorps anti-VIH se développent. Chez la moitié des patients environ, un syndrome grippal caractérise cette phase.
- Au cours de la **phase asymptomatique** de la maladie, aucun signe ou symptôme lié à l'infection par le VIH n'est visible. La durée de cette phase est très variable selon les patients : de moins d'un an jusqu'à plus d'une dizaine d'années. Au cours de cette phase, hors traitement antirétroviral, la charge virale augmente, les lymphocytes CD4 diminuent, affaiblissant ainsi le système immunitaire.
- Le **stade SIDA** (Syndrome d'Immuno-Déficience Acquise) est la forme la plus évoluée de l'infection par le VIH. Il survient chez des patients fortement immuno-déprimés, c'est-à-dire dont le niveau des lymphocytes CD4 est très faible (en dessous de 200 CD4/mm<sup>3</sup>) et dont le système de défense de l'organisme est particulièrement affaibli. A ce stade, apparaissent les infections, dites opportunistes, car elles profitent de l'état de faiblesse des défenses immunitaires pour se développer dans l'organisme. Parmi les plus fréquentes, on peut citer la toxoplasmose ou des atteintes pulmonaires graves, des tumeurs de type sarcome de Kaposi ou des infections virales. L'espérance de vie des patients à ce stade dépend, entre autre, des traitements et des infections opportunistes qui se développent.

Dans l'infection par le VIH (Virus d'Immuno-déficience Humaine), les traitements antirétroviraux (ARV) donnés sous la forme de multi-thérapies (connus sous le nom de HAART : Highly active antiretroviral therapy) ont montré leur efficacité pour :

- empêcher le VIH de progresser, en réduisant la charge virale (quantité de virus dans le sang) jusqu'à des taux indétectables (moins de 50 copies/ml) aussi longtemps que possible, et en limitant le développement de résistance aux médicaments ARV (c'est-à-dire lorsque le virus arrive à se multiplier en présence d'un traitement) ;
- restaurer et maintenir le fonctionnement normal du système immunitaire, en maintenant le taux de CD4 à des valeurs supérieures à 500 cellules/mm<sup>3</sup> ou en permettant d'augmenter ce taux s'il est trop faible ;

- améliorer la survie et la qualité de vie des patients infectés par le VIH, en diminuant le risque de complications et d'infections liées au VIH, et en réduisant les effets secondaires liés aux antirétroviraux.

**Illustration de l'évolution de l'infection à VIH au cours du temps**



Le nombre de traitements et le nombre de combinaisons possibles de ces traitements étant limités, il est crucial que le premier traitement donné aux patients soit optimisé en terme de puissance antivirale et de tolérance pour améliorer le pronostic vital. Malgré ces recommandations, des études épidémiologiques et sociologiques récentes ont montré qu'une part importante de malades arrêtaient leur traitement au cours de la première année, malgré des nouveaux traitements mieux tolérés. On pense que certains facteurs sociaux et médicaux pourraient influencer le niveau d'observance (bonne prise du traitement), et de ce fait, la probabilité d'arrêter le traitement.

Dans cette problématique, une étude statistique a été mise en place pour mettre en évidence les différentes causes d'arrêt de traitement et pour identifier les facteurs associés à la durabilité des premières lignes de traitements antirétroviraux.

## 2 Données

Les données que vous analyserez pour ce projet proviennent de la cohorte NADIS (New AIDS Data Information System) : une cohorte multicentrique alimentée par un dossier médical informatisé (DMI). Les données (socio-démographiques et cliniques) sont saisies par les soignants en temps réel lors des consultations ou des hospitalisations. Ce DMI est utilisé dans 30 centres hospitaliers français de la métropole, des Antilles Françaises et de la Guyane.

Ont été inclus, dans l'étude, 1136 patients infectés par le VIH, âgés d'au moins 18 ans, naïfs d'antirétroviraux et initiant une première ligne de HAART entre le 1er janvier 2000 et le 30 juin 2008 dans l'un des 8 sites utilisant NADIS® et fournissant des données de qualité contrôlée. Les variables retenues concernent l'état civil du patient, la maladie et son évolution, le traitement reçu et l'arrêt de ce traitement. La liste détaillée des variables est donnée ci-après.

## 3 Objectifs du projet

L'objectif général de ce projet est d'étudier les arrêts de traitement survenant au cours du suivi, leurs motifs et les facteurs pouvant favoriser ou empêcher leur survenue. Différents critères concernant l'arrêt du traitement peuvent être considérés dans l'analyse statistique : arrêt pour toutes causes (présence/absence), arrêt pour cause d'intolérance (présence/absence), arrêt pour échec immuno-virologique (présence/absence), .... L'arrêt est défini comme toute interruption d'un ou de plusieurs médicaments de la 1ère ligne, sans tenir compte des variations de posologies.

Dans cette analyse, vous pourrez utiliser les méthodes d'analyse statistique que vous avez déjà étudiées en L3 et M1, et mettre en application de nouvelles méthodes que nous introduirons au cours du BE selon les besoins. Plusieurs analyses sont possibles. En effet, on dispose ici de données longitudinales et il existe de nombreux façons de les traiter et de les représenter.

Ce projet s'appuie sur des données réelles. La présence de valeurs manquantes est possible et vous devez les prendre en considération dans votre travail. De plus, vous pouvez (et devrez) également créer de nouvelles variables qui seraient appropriées pour votre analyse. Tout cela devra être expliqué dans votre rapport.

Avant de commencer votre travail d'analyse statistique, établissez un plan précis de vos analyses et des méthodes à utiliser (selon la nature des variables à utiliser).

On vous propose un plan d'analyse que vous pourrez ajuster et compléter selon vos idées :

### 1. Description de la population étudiée

Décrire les caractéristiques des patients (critères socio-démographiques, maladie, traitement). On s'intéressera en particulier au critère principal de cette étude, à savoir la présence d'un arrêt et les différents motifs. On pourra également penser à une représentation graphique des arrêts survenant au cours du temps (globalement et selon les différents types).

## 2. Traitement des données manquantes

Comme vous le verrez dans les premières analyses descriptives, vous serez confrontés ici à la présence (parfois importante) de données manquantes (ou éventuellement, aberrantes). Elles devront être prises en compte dans cette étude statistique (suppression des individus, nouvelle modalité, remplacement, ...). A la fin de cette étape, il est impératif que votre base de données soit figée et que toutes les analyses que vous menerez par la suite soient réalisées sur le même nombre d'observations.

## 3. Relation entre l'arrêt du traitement et les caractéristiques des patients

Des analyses bivariées et des modélisations pourront être mises en œuvre pour mettre en évidence les caractéristiques pouvant être liées à l'arrêt du traitement ou à un type d'arrêt de traitement.

## 4. Modélisation longitudinale par des méthodes d'analyse de survie

Dans cette étude, l'évènement est l'arrêt du traitement (toutes causes ou par cause) : s'il ne survient pas au cours du suivi, on sait que le délai de survenue de l'évènement est supérieur à la durée du suivi ; on dit que l'évènement est censuré. Des méthodes existent permettant d'analyser des données censurées en prenant en compte toute l'information disponible.

La première difficulté de cette partie résidera dans la définition de la variable arrêt/censure et du délai correspondant.

Dans un deuxième temps, vous pourrez étudier l'influence éventuelle des caractéristiques des patients sur la survenue des arrêts au cours du temps.

# 4 Liste des variables

Les données sont disponibles dans le fichier `base_sid2013.csv` contenant les variables suivantes (espacées d'une virgule) :

- `num` : identifiant,
- `age` : âge à l'initiation du traitement, en classes
  - 0 = [18 ; 30[
  - 1 = [30 ; 40[
  - 2 = [40 ; 50[
  - 3 = [50 ; 60[
  - 4 = supérieur à 60
- `sexe` : sexe
  - 1 = homme
  - 0 = femme
- `observ` : consultation d'observance
  - 0 = non
  - 1 = oui

Remarque : l'observance est utilisée pour mesurer si un patient respecte la posologie de son traitement selon la prescription médicale.

- **sida** : stade sida
  - 0 = non
  - 1 = oui
- **conta** : mode de contamination par le VIH
  - 0 = hétérosexuel
  - 1 = homo/bisexuel
  - 2 = toxico par voie intraveineuse
  - 3 = autres (transfusé, hémophile, ...)
- **typetrt** : type de traitement reçu
  - 0 = 2IN+1IP
  - 1 = 2IN+INN,
  - 2 = autres (mono, bi, quadrithérapie, ...)

Remarque : les multithérapies associent plusieurs types de médicaments : les INN (inhibiteurs non-nucléosidiques), les IN (inhibiteurs nucléosidiques) et les IP (inhibiteurs de protéase) qui se différencient selon leurs modes d'action sur le développement du virus.

- **cd4b1** : CD4 à l'initiation du traitement
  - 0 = moins de 200
  - 1 = entre 200 et 350
  - 2 = 350 et plus
- **cd4b2** : CD4 à l'arrêt du traitement ou à la fin du suivi sans arrêt
  - 0 = moins de 200
  - 1 = entre 200 et 350
  - 2 = 350 et plus

Remarque : il se peut que les deux valeurs de CD4 (et de charges virales données ci-après) soient égales dans le cas où l'arrêt du traitement se produit très rapidement après le début du traitement.

- **cvb1** : Charge virale à l'initiation du traitement
  - 0 = entre 1.7 et 4
  - 1 = entre 4 et 5
  - 2 = 5 et plus
- **cvb2** : Charge virale à l'arrêt du traitement ou à la fin du suivi sans arrêt
  - 0 = moins de 1.7
  - 1 = entre 1.7 et 4
  - 2 = 4 et plus

Remarque : Les classes de cvb1 et cvb2 ne sont pas les mêmes. En effet, à l'initiation au traitement, la charge virale est presque toujours détectable (supérieure à 1.7) alors qu'elle peut devenir indétectable au cours du traitement (inférieure à 1.7).

- **d\_debtrt** : Date de début du traitement

- **debtrtc** : année de début du traitement en deux classes

- 0 = 2004 et après

- 1 = avant 2004

Remarque : L'année 2004 marque un changement dans les multithérapies (prise en charge, tolérance, efficacité, ...).

- **d\_sero** : Date de séropositivité (marque le début connu de l'infection).

- **d\_point** : Date de point fixée au 30 juin 2008

Remarque : on ne considère pas les événements survenus au delà de cette date du 30 juin 2008 ; si un arrêt de traitement est intervenu après cette date, le suivi du patient est considéré comme censuré à la date de point et l'information sur l'arrêt du traitement n'est pas prise en compte dans l'analyse statistique.

- **d\_arret** : Date d'arrêt du traitement

Remarque : cette variable n'est pas renseignée si le patient n'a pas arrêté son traitement au cours du suivi. Elle peut aussi être renseignée, mais non prise en compte dans les analyses, si l'arrêt est survenu après la date de point.

- **motif** : motif d'arrêt du traitement

- intolérance/toxicité

- échec thérapeutique

- autres (simplification du traitement, problème d'observance, ...)

Remarque : si cette variable n'est pas renseignée, cela signifie que le patient n'a pas arrêté son traitement au cours de l'étude, c'est-à-dire avant le 30 juin 2008. Il se peut que le patient ait arrêté son traitement après le 30 juin 2008 et que cette variable soit renseignée ainsi que la date d'arrêt. Toutefois, la date de point de l'étude étant fixée au 30 juin 2008, les arrêts survenus après cette date ne doivent pas être pris en compte dans l'analyse : ce patient est alors considéré comme n'ayant pas arrêté son traitement au cours de l'étude.