

# Phase IV, pharmacovigilance

-

## Identification des effets indésirables médicamenteux

- I. Détection des effets indésirables médicamenteux (EIM)
- II. Systèmes de règles de décision
- III. Data mining et arbres de décision
- IV. Exemple de détection automatisée des EIM par arbres de décision

# I. Détection des effets indésirables des médicaments (EIM)

- A. Définition des EIM
- B. Différents cadres d'étude
- C. Synthèse

## *Détection des EIM*

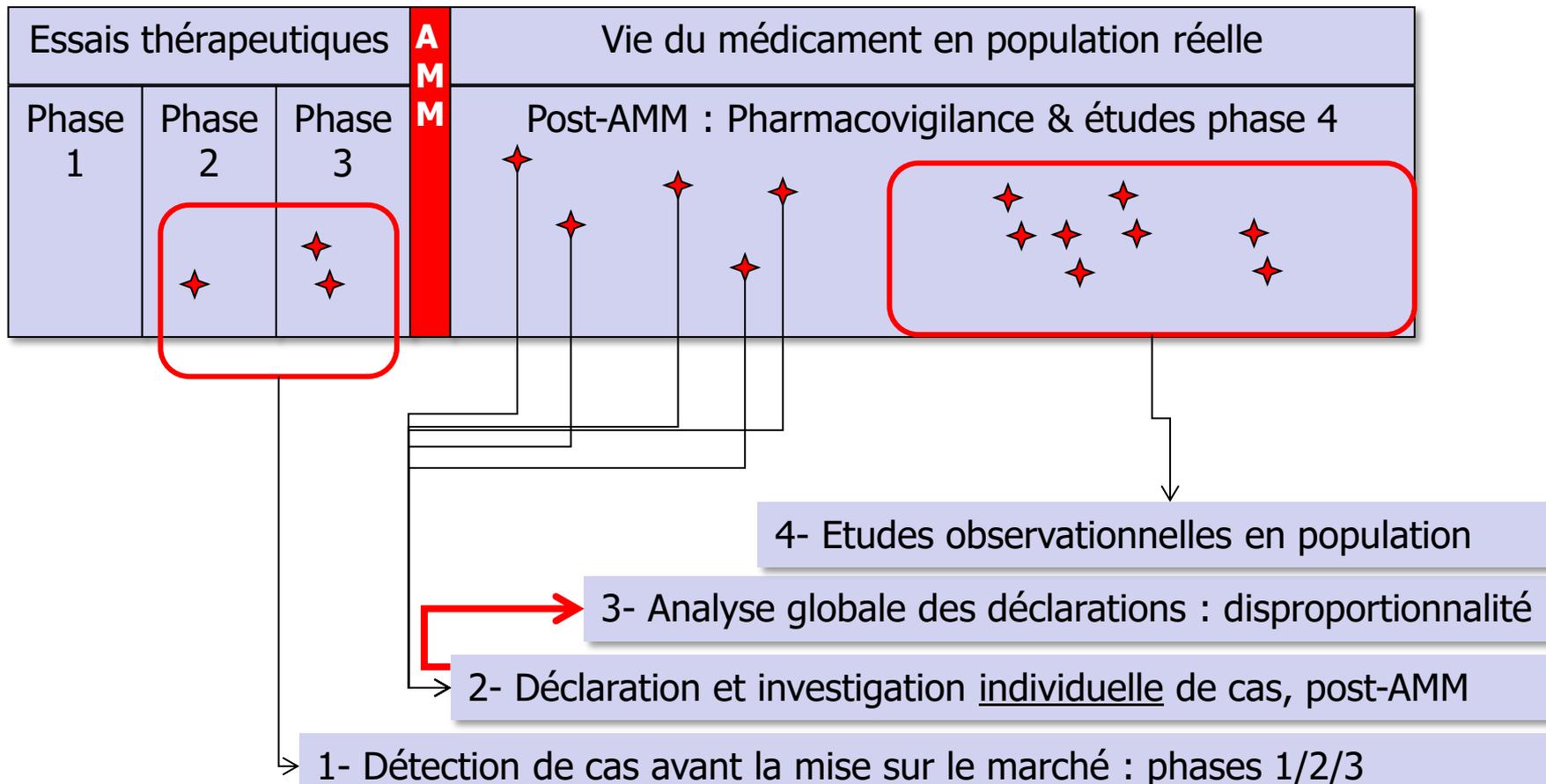
# Effets indésirables des médicaments

- Plusieurs définitions, nuances assez complexes...
- Définition utilisée ici :
  - Réaction nocive et non voulue,
  - Survenant aux doses normalement utilisées ( $\neq$  erreur d'administration) chez l'homme,
  - Liée à l'utilisation du médicament (prise ou arrêt)
- Estimation épidémiologique :
  - 15 000 morts par an en France
  - Principalement chez les personnes âgées, sous anticoagulants ou insuline
  - Coût économique majeur (hospitalisations, rallongement des durées de séjour)

# Détection des EIM

## Méthodes d'étude des EIM

- En fonction de la vie d'un médicament :



# 1- dans les phases 2 et 3

- Définition :
  - Phases 2 et 3 (cf. cours Pr Alain Duhamel) : étude du médicament sur des échantillons de patients volontaires
- Avantages :
  - Avant la mise sur le marché
  - Protocole idéal : essai randomisé (double aveugle) contrôlé (versus placebo ou traitement de référence). Donc inclut des patients non-traités par ce médicament ( $\neq$  phase 1)
  - Recueil d'EIM exhaustif, bien documenté
- Inconvénients :
  - Peu de patients
  - Profils de patients plus simples que dans la population réelle : âge limité, pas de femme enceinte, peu de pathologies associées, profil pathologique « pur », un minimum de traitements concomitants
  - Contexte de prescription trop parfait : praticiens très formés (pas de mésusage), surveillance rapprochée

# Détection des EIM

## 2- déclaration individuelle des cas

- Définition :
  - Déclaration obligatoire des cas observés par les professionnels de santé
  - Analyse du cas individuellement en utilisant des méthodes d'analyse d'imputabilité (=grille de lecture et évaluation du cas)
- Avantages :
  - En population réelle (vrai patients, vrais usages, vrais prescripteurs)
  - Potentiellement beaucoup de patients... si les médecins identifient le cas comme étant un EIM, et un EIM rare ou grave
  - Recueil exhaustif de chaque cas déclaré
  - Déclaration est effective précocement après l'AMM
- Inconvénients :
  - Très peu d'EIM sont déclarés : à peine 5% des EIM graves selon différentes études ! inversement, effets de mode => impossible d'estimer l'incidence
  - EIM parfois pas envisagé par les médecins => il manque certainement des situations
  - Protocole très simple : on revoit UN cas exposé et avec événement (aucun témoin, aucun non-exposé, pas de statistiques)
  - Niveau de preuve faible

## 3- analyses de disproportionnalité

- Définition :
  - Lors de la déclaration d'un EIM, TOUS les médicaments du patient sont déclarés, qu'ils soient en relation ou non avec l'EIM
  - Tous ces EIM constituent une base de données : que des EIM, mais de natures différentes
  - Identification de médicaments statistiquement associés avec des types d'EIM (disproportionnalité : non-indépendance entre l'exposition et le type d'EIM, même si on n'a que des EIM). Cas-témoin avec des témoins un peu particuliers.
  - (Pour mémoire : études parfois appelées « cas – non-cas »)
- Avantages :
  - En population réelle (vrai patients, vrais usages, vrais prescripteurs)
  - Potentiellement beaucoup de patients... (malgré sous-déclaration)
  - Recueil exhaustif de chaque cas déclaré
- Inconvénients :
  - Très peu d'EIM sont déclarés
  - Témoins atypiques : ce sont des EIM de nature différente, pas des cas sans EIM
  - Impossible d'estimer une incidence

# 4- études observationnelles en population

### ■ 4a- Etudes épidémiologiques classiques

- Définition :
  - Véritables études épidémiologiques : cas/témoin, exposé/non-exposé ou cohorte (cf. cours de PACES)
  - Cas = EIM étudié, témoin = sans EIM
  - Exposés = avec le médicament d'intérêt, non-exposés = sans le médicament d'intérêt
- Avantages :
  - En population réelle (vrai patients, vrais usages, vrais prescripteurs)
  - Cas bien documenté en particulier pour les études prospectives
  - Niveau de preuve plus élevé
  - Permet enfin d'estimer une incidence
- Inconvénients :
  - Tardif, cher

### ■ 4b- Etudes de réutilisation de données (data reuse)

- Définition :
  - Réutilisation de données de prescription (bases assurance maladie), d'hospitalisations (PMSI) avec actes et diagnostics, de décès
  - Véritables cohortes historiques
- Avantages :
  - Fort niveau de preuve, effectif très élevé donc forte puissance (parfois exhaustives, nationales)
  - Peu onéreuses, résultats rapides
- Inconvénients :
  - Manque d'information détaillée sur les cas (absence des symptômes moins sévères, etc.), pas d'évaluation de l'imputabilité
  - Longtemps après l'AMM

# Détection des EIM

## Tableau récapitulatif

	1- Phases 2 et 3	2- Déclaration individuelle de cas	3- Analyses de disproportionnalité	4a- Etudes épidémio. en population	4b- Etudes de réutilisation de données
Chronologie	Avant AMM	Post-AMM immédiat	Post-AMM	Post-AMM tardif	Post-AMM tardif
Protocole	Essai randomisé contrôlé	Analyse d'un patient exposé et avec événement	Analyse des bases contenant les déclarations d'EIM	Véritables études épidémiologiques	Cohortes historiques sur BDD médico-administratives
Patients, contexte, médecins	Trop parfaits	Réalistes	Réalistes	Réalistes	Réalistes mais informations partielles
Détection des EIM	Exhaustive, bien documentée.	Très peu sensible, mais bien documentée	Très peu sensible, mais bien documentée	Très bonne	Variable selon le type d'EIM
Nombre d'EIM vus	Trop faible	Potentiellement élevé, mais effet de mode / focus	Potentiellement élevé, mais effet de mode / focus	Elevé (selon le budget de l'étude)	Souvent très élevé
Estimation incidence	Imprécise	Impossible	Impossible	Excellente	Bonne
Imputabilité	Etudiée	Etudiée	Ignorée	Variable	Ignorée
Asso statistique	Oui (RR)	Non	Oui (OR)	Oui (OR/RR)	Oui (RR)
Niveau preuve	1 😊😊	4 😞	3 😞	2-3 😊	2-3 😊

## Détection des EIM

# Population concernée, si on s'intéresse à un couple médicament-EIM

- 1- Détection de cas avant la mise sur le marché : phases 2 et 3

	EIM	Autre EIM	Pas d'EIM
Méd présent	●		●
Méd absent	●		●

- 2- Déclaration et investigation individuelle de cas, post-AMM

	EIM	Autre EIM	Pas d'EIM
Méd présent	◆ 1 cas		
Méd absent			

- 3- Analyse globale des déclarations : disproportionnalité

	EIM	Autre EIM	Pas d'EIM
Méd présent	●	●	
Méd absent	●	●	

- 4- Etudes observationnelles en population

	EIM	Autre EIM	Pas d'EIM
Méd présent	●●	●●	●●
Méd absent	●●	●●	●●

# II. Systèmes de règles de décision

- A. Définition, formalisme
- B. Métriques pour une règle isolée
- C. Systèmes à plusieurs règles
- D. Métriques pour un système de règles

# Systemes de règles de décision

## A. Définition, formalisme

- Règles de décision :
  - ensemble de conditions qui, réunies, produisent un effet avec une certaine probabilité  
 $C_1 \& C_2 \& \dots \& C_k \rightarrow E$
  - Expression usuelle de la connaissance en médecine
  - Exemple : Héparine & âge>70 → hémorragie  
(parfois seulement !)
- Notations :
  - On utilisera indifféremment  $C_i$  pour un événement, ou pour l'ensemble correspondant :  
 $C_1 \cap C_2 \cap \dots \cap C_k \rightarrow E$
  - Pour simplifier, soit  $C$  le fait que les conditions soient réunies  
 $C = C_1 \cap C_2 \cap \dots \cap C_k$   
alors la règle peut s'écrire  
 $C \rightarrow E$

# *Systemes de règles de décision*

## A. Définition, formalisme

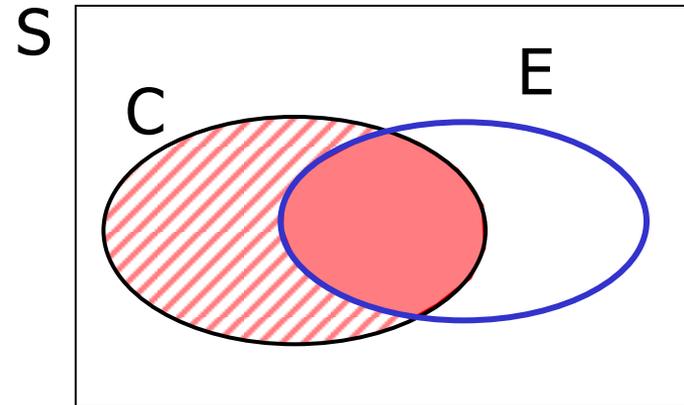
- Éléments de la règle :
  - Conditions = événements binaires éventuellement issus de variables non binaires :
    - Variable d'emblée binaire : « insuffisance rénale {0/1} »
    - Depuis une variable quantitative : « âge > 70 ans »
    - Depuis une variable qualitative : « couleur = "brun" »
  - Conditions reliées par l'opérateur « et »
  - Effet = événement binaire (obtenu de la même manière)
- Métriques : la confiance et le support

# Systemes de règles de décision

## B. Métriques pour une règle isolée

### Confiance :

- Probabilité que l'effet se produise, sachant que les conditions sont réunies
- C'est la VPP des conditions
- Confiance  
=  $P(E | C)$   
=  $\#(C \cap E) / \# C$
- => *prédit un risque dans un groupe de patients*

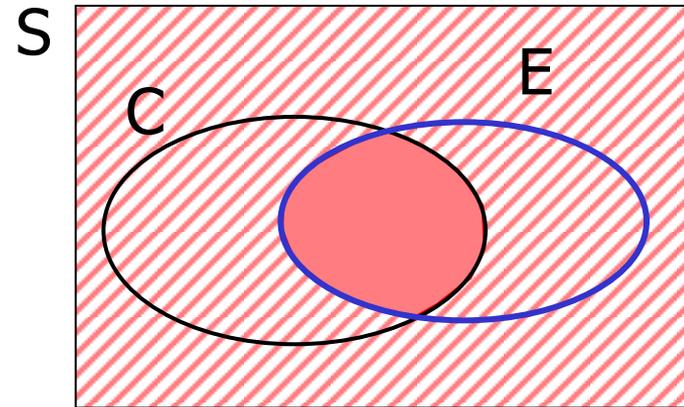


# Systemes de règles de décisions

## B. Métriques pour une règle isolée

### Support :

- Probabilité que les conditions et l'effet se produisent
- Support  
 $= P(E \cap C)$   
 $= \#(C \cap E) / \# S$
- $\Rightarrow$  *renseigne sur la portée de la règle : est-ce une niche ?*



# *Systemes de règles de décision*

## B. Métriques pour une règle isolée

~~Sensibilité,~~

~~Spécificité,~~

~~VPN :~~

- Théoriquement calculables
- Mais aucun intérêt dans le cas d'une règle isolée
- Un individu non détecté par la règle d'intérêt est peut-être détecté par une autre règle...

# Systemes de regles de decision

## C. Systemes à plusieurs regles

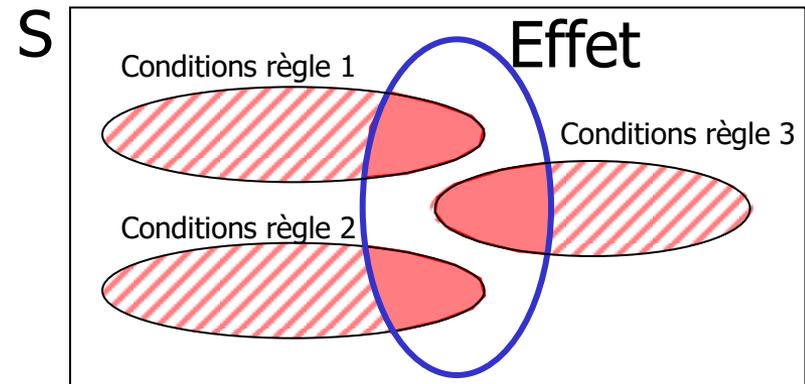
- Systeme où plusieurs regles prédisent le même effet.
- Exemple de système de regles avec redondance :
  - *Médicament A* → *hémorragie* (*confiance=20%*)
  - *Médicament B* → *hémorragie* (*confiance=60%*)
  - ... or certains patient ont médicament A et médicament B  
=> Pour un patient donné, quelle confiance choisir ?
- Exemple de système de regles sans redondance :
  - *Médicament A* → *hémorragie*
  - *Pas de médicament A & Médicament B* → *hémorragie*
  - Personne ne remplit les conditions des deux regles en même temps  
=> une seule confiance pour un patient, mais regles plus longues et de support généralement plus faible

# Systemes de regles de decision

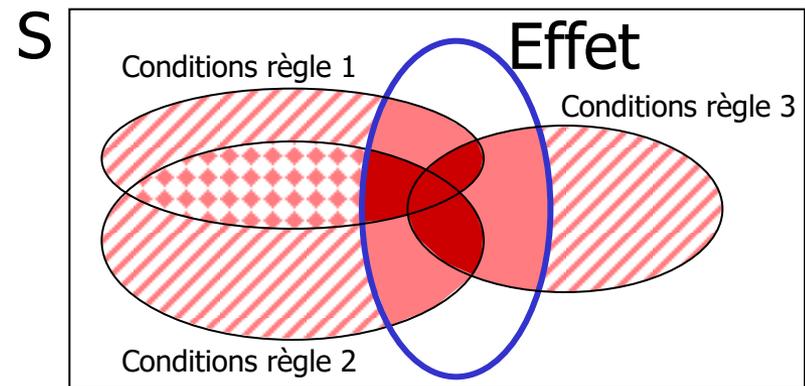
## D. Métriques pour un système de règles

### La « précision » :

- VPP globale du système de règles
- Probabilité globale d'observer l'effet lorsque les conditions d'au moins une règle sont réunies
- NB : si système non redondant, moyenne pondérée des confiances des règles



*Système non redondant*



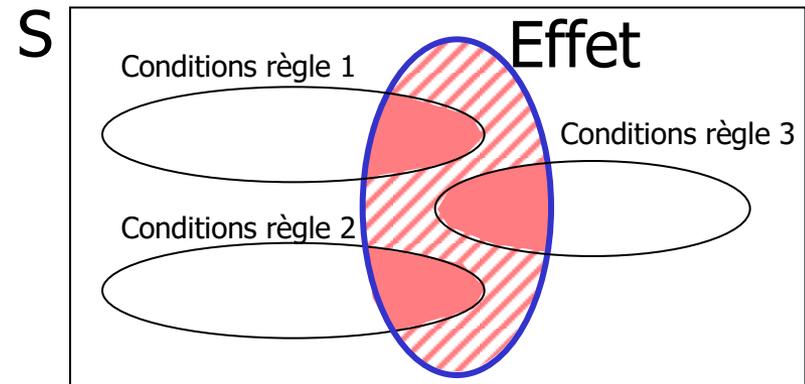
*Système redondant*

# Systemes de règles de décision

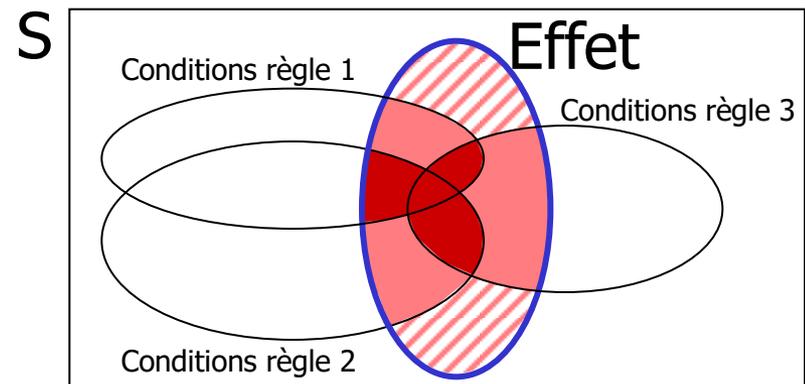
## D. Métriques pour un système de règles

### Le « rappel » :

- Sensibilité globale du système
- Probabilité qu'un effet soit prédit par au moins une règle
- NB : si système non redondant, somme des sensibilités des règles



*Systeme non redondant*



*Systeme redondant*

# III. Data mining et arbres de décision

- A. Définition du data mining
- B. Méthodes
- C. Les arbres de décision

# *Data Mining et arbres de décision*

## A. Définition du data mining

- = fouille statistique de données
- « *Mise en évidence de connaissances jusqu'alors inconnues dans des bases de données de grande dimension, à l'aide de méthodes dérivées des statistiques, de la gestion de données et de l'intelligence artificielle* »
- Grande dimension (big data) :
  - Beaucoup de lignes (patients)
  - Beaucoup de colonnes (variables)
  - Beaucoup de cardinalité (tables et relations)

# Data Mining et arbres de décision

## B. Méthodes

- Data mining non supervisé :
  - recherche de motifs et associations
  - Ex : règles d'associations (problème du « panier de la ménagère »)
- Data mining supervisé :
  - Objectifs :
    - Rétrospectif : expliquer une variable  $Y$  (à expliquer) par les autres variables  $X_i$  (explicatives)
    - Prospectif : prédire la valeur inconnue de cette variable  $Y$  par les autres variables  $X_i$
  - Ex : **arbres de décision**
- Méthodes utilisables pour :
  - 3- Analyse globale des déclarations : disproportionnalité
  - 4- Etudes observationnelles en population
  - **!/! ce ne sont pas les méthodes classiques de référence, mais des méthodes exploratoires que nous souhaitons vous présenter.**

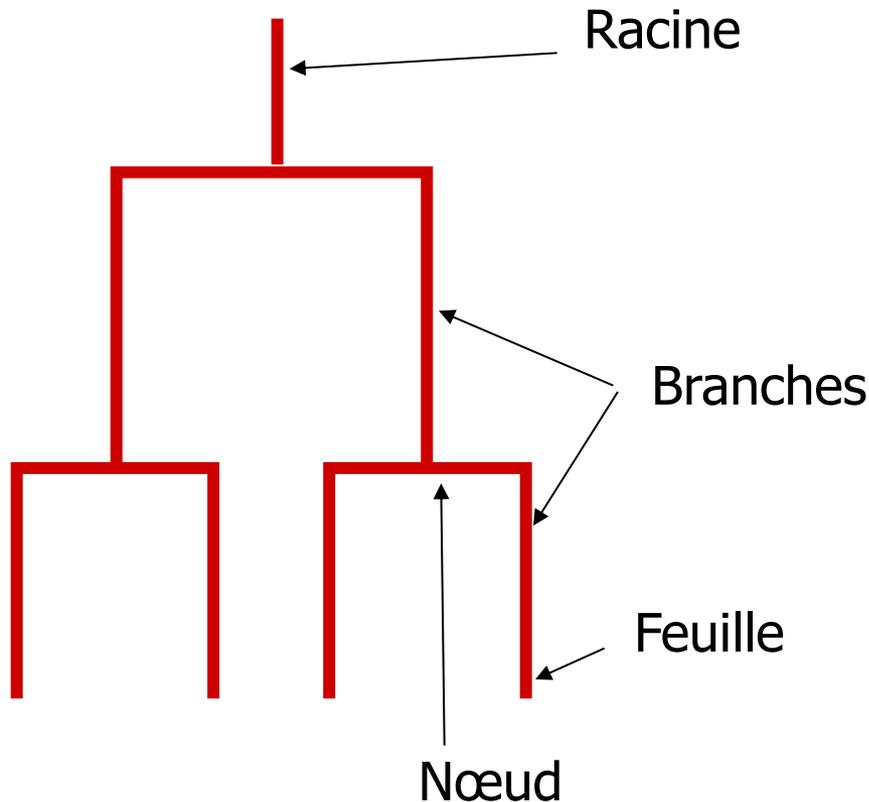
# *Data Mining et arbres de décision*

## C. Les arbres de décision

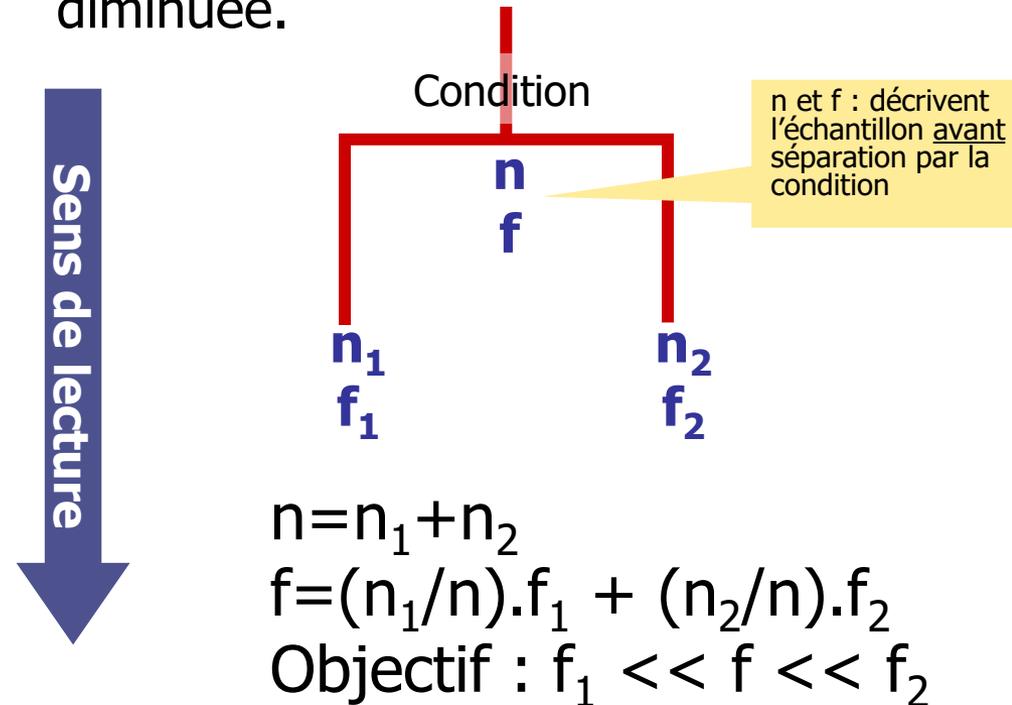
- Méthode de data mining supervisé
- Produit des règles
- Variable à expliquer  $Y$  :
  - variable binaire principalement
  - (mais il existe d'autres types d'arbres...)
- Variables explicatives  $X_i$  :
  - utilise tous les types de variables sous forme binaire (cf. précédemment)
  - peut mélanger des variables de types différents

# Data Mining et arbres de décision

## C. Les arbres de décision binaires

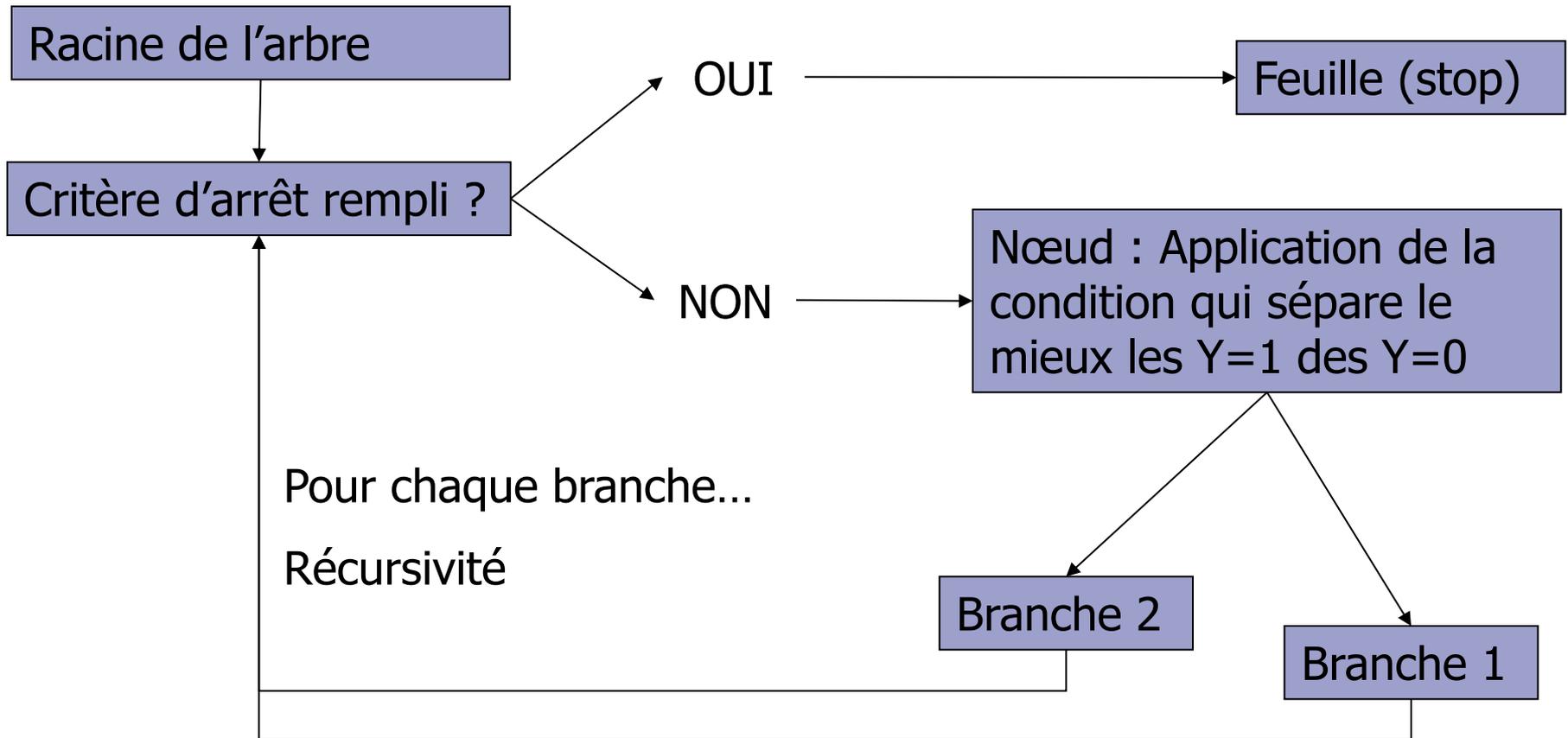


Unité de base : le nœud sépare une branche en deux branches à l'aide d'une condition. Ces deux branches : une avec  $f$  augmentée, l'autre avec  $f$  diminuée.



# Data Mining et arbres de décision

## C. Les arbres de décision binaires

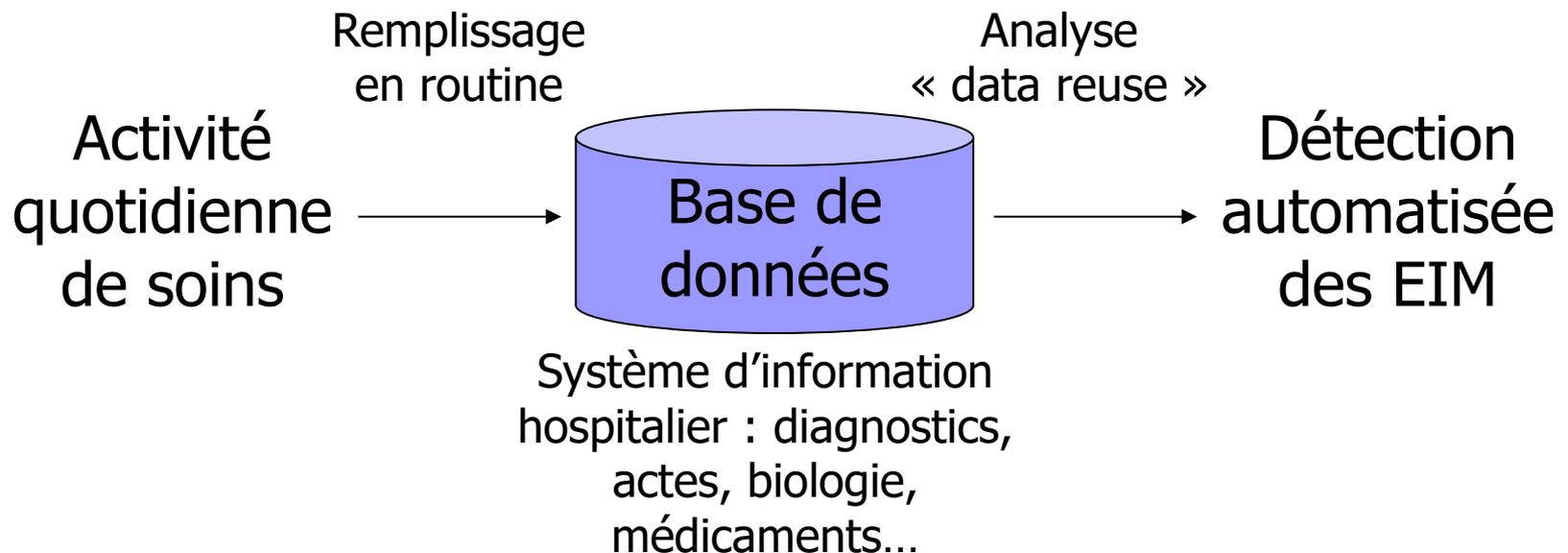


# IV. Détection automatisée des EIM par arbres de décision

- A. Objectifs, schéma général
- B. Approche par l'exemple
- C. Interprétation de l'exemple
- D. Epilogue
- E. Réutilisation de données

## A. Objectifs, schéma général

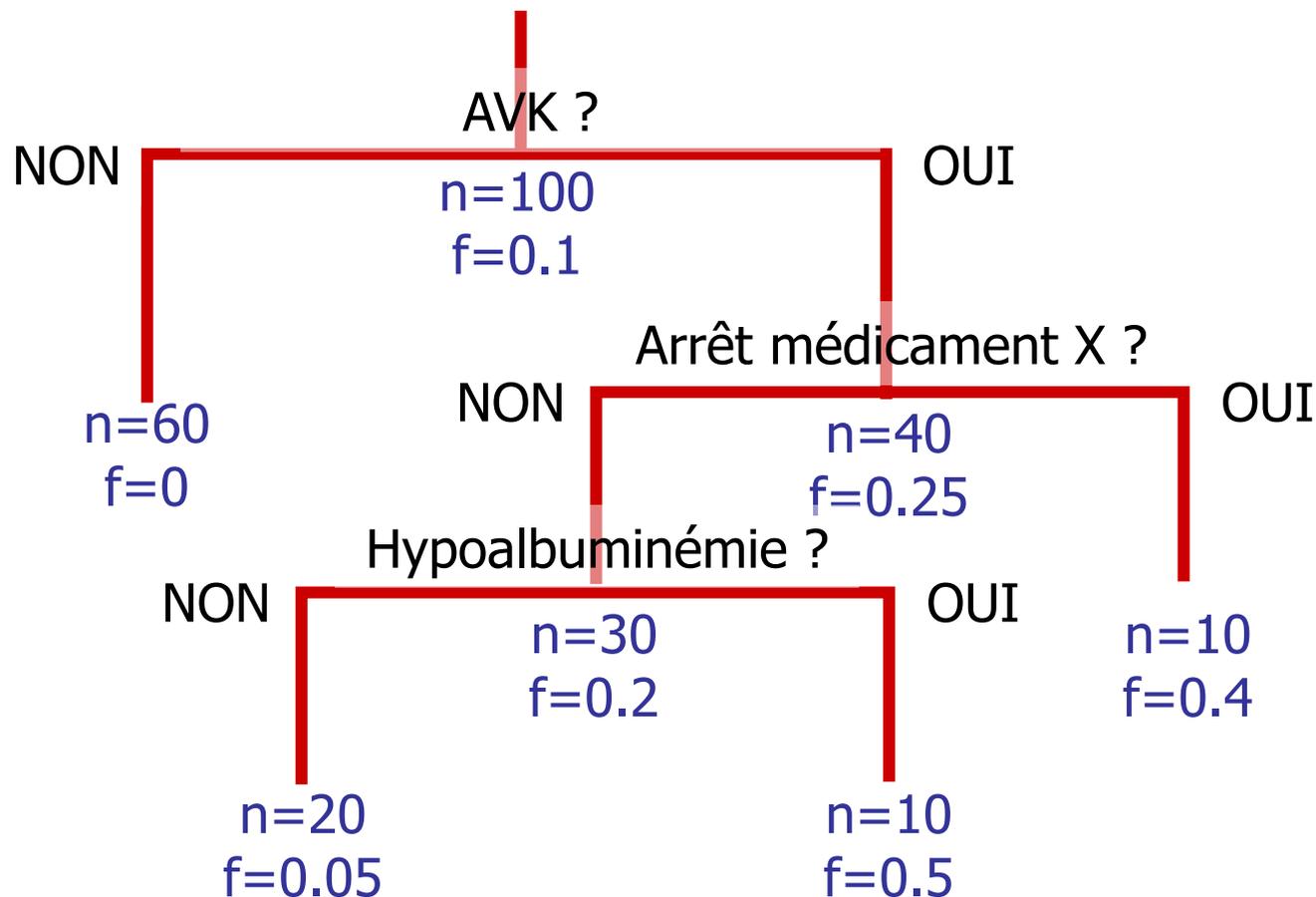
- Découvrir automatiquement des motifs de conditions (incluant des médicaments) fréquemment associés à des événements nocifs
- Par « data reuse » : réutilisation de données



## B. Approche par l'exemple

- Situation médicale :
  - Les antivitamine K (AVK) sont des anticoagulants, indiqués en prévention des thromboses
  - Activité biologique suivie en mesurant l'INR (international normalized ratio) sur des prélèvements sanguins
  - Effet indésirable fréquent :  
*surdosage en AVK → élévation de l'INR → hémorragie*
  - Nombreuses interactions médicamenteuses :  
à posologie constante, l'activité biologique varie fortement
- Echantillon : 100 séjours hospitaliers
- Variable à expliquer Y :
  - Y=1 (10%) : le patient présente un INR>5 (risque hémorragique)
  - Y=0 (90%) : pas d'anomalie
- Variables explicatives  $X_i$  :
  - $X_1$  à  $X_{500}$  : 500 variables binaires, portant sur les médicaments reçus et résultats de biologie

# B. Approche par l'exemple



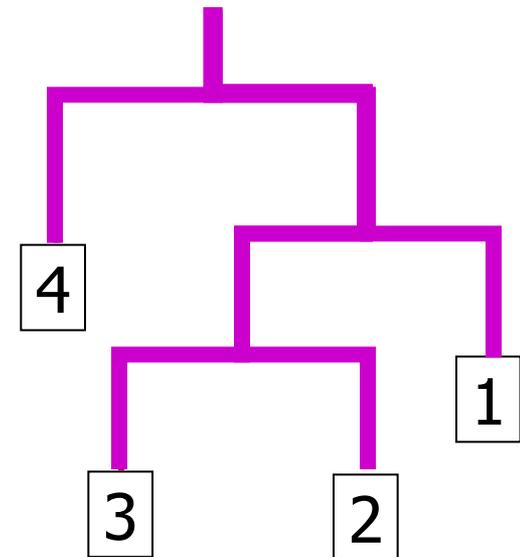
## B. Approche par l'exemple

4 feuilles => système non redondant de 4 règles

Enoncé des 4 règles de droite à gauche

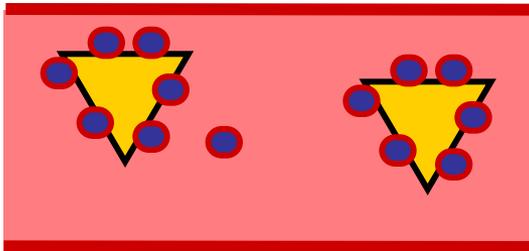
$P = \text{Probabilité}(Y=1) = \text{confiance}$ , estimée dans l'échantillon :

1. AVK & arrêt médicament X  $\rightarrow P=0.4$
2. AVK & pas d'arrêt médicament X & hypoalbuminémie  $\rightarrow P=0.5$
3. AVK & pas d'arrêt médicament X & pas d'hypoalbuminémie  $\rightarrow P=0.05$
4. Pas d'AVK  $\rightarrow P=0$



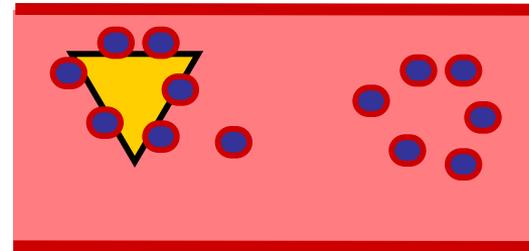
## C. Interprétation de l'exemple

Albumine = une protéine plasmatique. L'AVK se lie à l'albumine, mais seule la fraction non liée est biologiquement active.



### Etat de base :

99% des AVK sont liés à l'albumine.  
Seuls 1% sont biologiquement actifs.  
La dose administrée en tient compte.



### Hypoalbuminémie :

diminution de la fraction liée,  
augmentation de la fraction libre  
=> à dose identique, activité biologique accrue  
=> surdosage alors que la dose administrée est  
« normale »

## D. Epilogue sur la méthode

- Arbres de décision : une méthode de *data mining* supervisé parmi d'autres. Détecteur d'interactions : effets conditionnels des covariables.
- Méthode CHAID (ne pas retenir le nom) :
  - Utilisée dans cet exemple, la plus ancienne
  - CHi-squared Automatic Interaction Detector
  - Prédit un Y binaire, avec des  $X_i$  de tous types qui sont binarisées
- Critère de choix des nœuds dans CHAID :
  - le nœud s'appuie sur la variable  $X_i$  la plus fortement associée à Y par un test du  $\text{Chi}^2$  ( $\Leftrightarrow$  petit p le plus faible)
- Critères d'arrêt dans CHAID : au cours de la pousse de l'arbre
  - Branche « pure » :  $f=0$  ou  $f=1$
  - OU Toutes les variables  $X_i$  sont constantes dans la branche
  - OU Aucun  $\text{Chi}^2$  significatif
- ➔ Il existe plusieurs autres méthodes d'arbres de décision

## D. Epilogue sur la méthode

- Avantages des arbres de décision :
  - Construction facile, paramétrable
  - Natures variées de  $Y$
  - Natures variées de  $X_i$ , utilisables en même temps
  - Fonctionne en grande dimension (nombreuses lignes, nombreuses colonnes) sans difficulté
  - Arbre facile à lire, règles faciles à utiliser
  - Règles : systèmes de règles non redondants
- Inconvénients des arbres de décision :
  - Risque de sur-ajustement (associations fortuites)
  - Instabilité des arbres
  - Moins performants que d'autres méthodes
  - Règles utiles à la prédiction mais pas forcément théoriquement valides (ex : variables explicatives corrélées entre elles)

# E. Réutilisation des données du dossier patient électronique

- Concept de « data reuse » (réutilisation des données collectées en routine), cf. banque, assurance, grande distribution...
- Données toujours disponibles :
  - Programme de médicalisation des systèmes d'information, PMSI : démographie, diagnostics, actes
- Données souvent informatisées :
  - Biologie
  - Courriers en texte libre
  - Médicaments administrés
- Dans le cas ci-après, quels éléments traceurs permettent de détecter (sans certitude) un surdosage en AVK ?

## Dossier administratif

Femme de 88 ans

## Diagnostics

I10 Hypertension artérielle  
Z8671 Antécédents personnels de cardiopathies ischémiques

## Actes

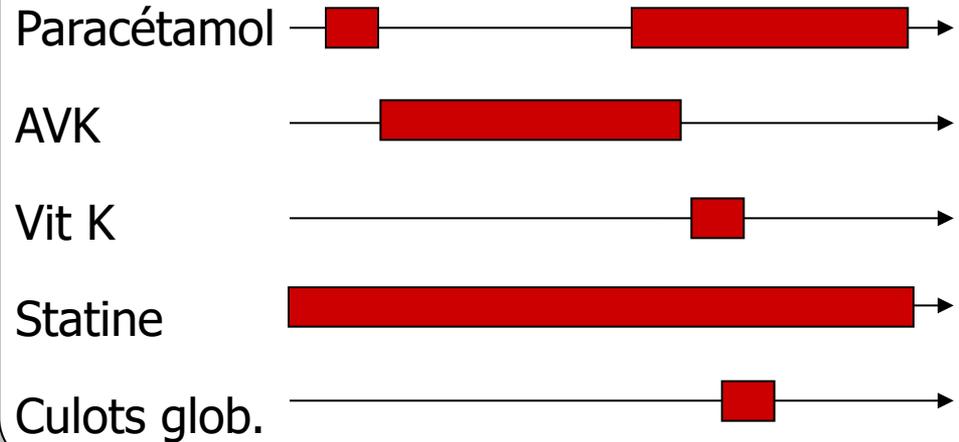
ZCQK002 Radio de l'abdomen sans préparation  
FELF001 Transfusion de concentré de globules rouges

## Courriers

Lettre de sortie

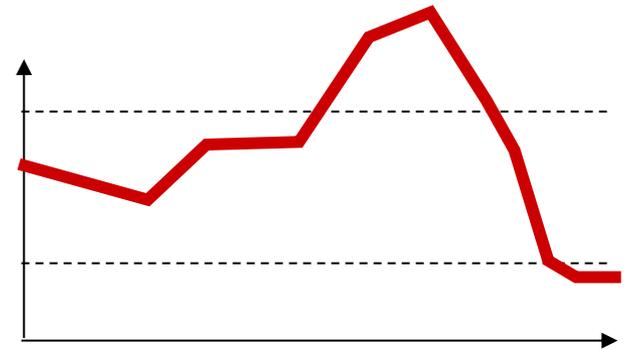
Compte-rendu opératoire

## Médicaments



## Biologie

INR



Hémo-globine

