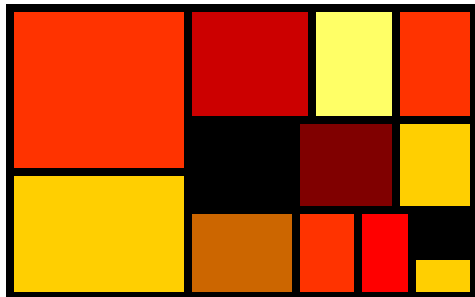
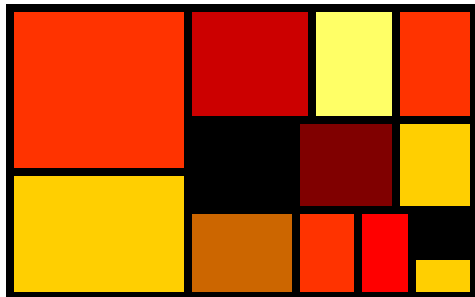


Indicateurs de qualité des soins et dossier patient électronique



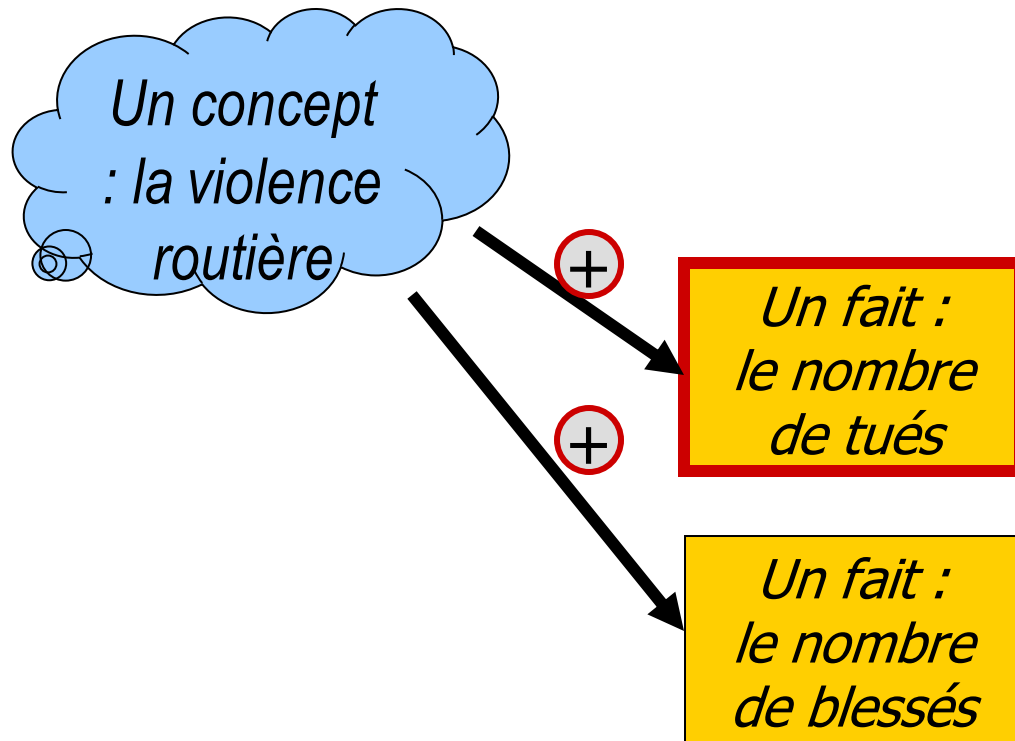
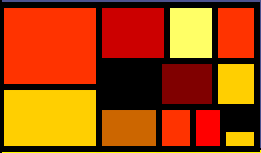
- I. Exemples introductifs
- II. Indicateurs nationaux, IPAQSS
- III. Dossier patient électronique
- IV. Calculer des indicateurs dans le dossier électronique
- V. Débogage d'indicateurs par Arbres de Décision
- VI. Certains calculs plus complexes...

Exemples introductifs



- I. Les indicateurs pour :
- II. mesurer quelque chose
- III. faire le point
- IV. évaluer et suivre une action
- V. ...faire croire !

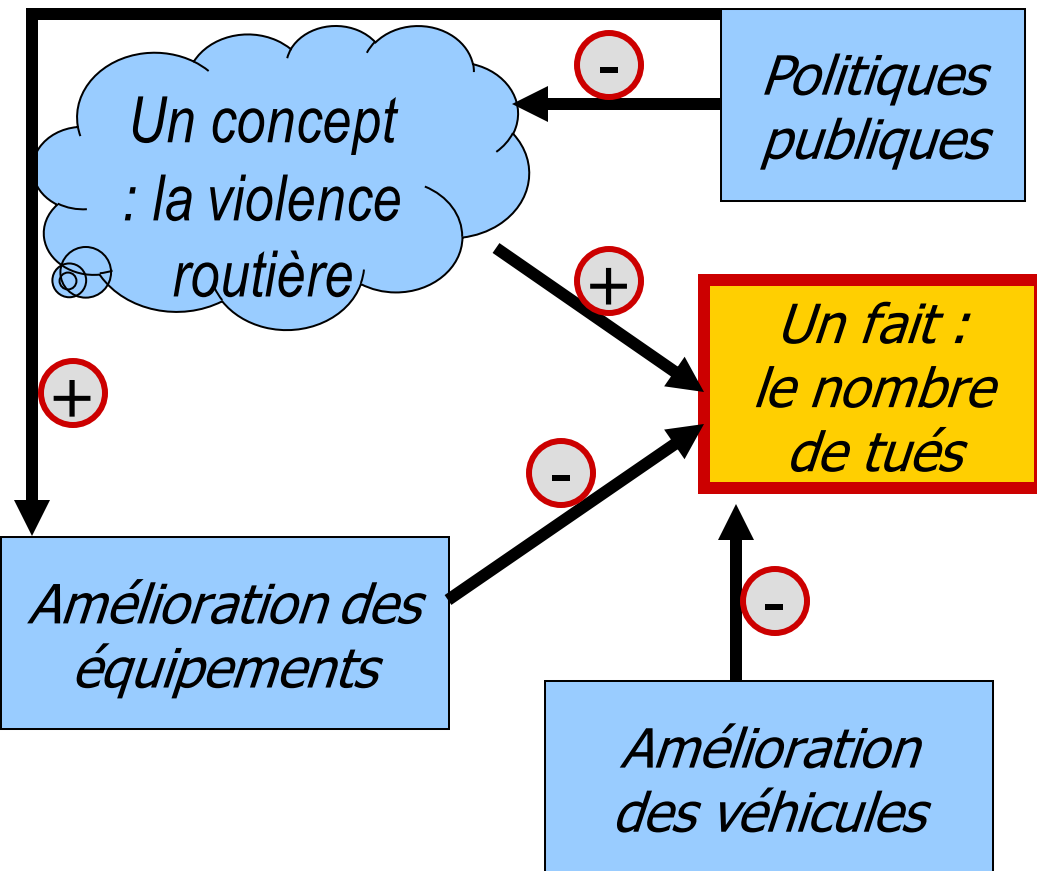
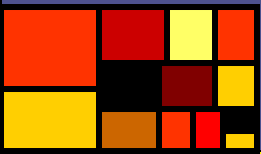
Le nombre de tués sur la route (1)



Approche statique :

- Pourquoi faire le point : quel phénomène sous-jacent l'indicateur mesure-t-il ?
- Le nombre de tués est-il la seule manifestation du phénomène ?
- Si non quel est son intérêt ?

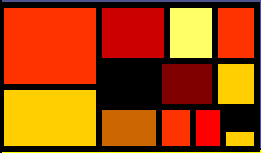
Le nombre de tués sur la route (2)



Approche dynamique :

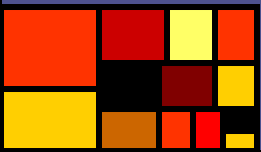
- Quelle est l'utilité de suivre cet indicateur dans le temps ?
- Sensibilité : un effet des politiques publiques sera-t-il visible ?
- Spécificité : les politiques publiques sont-elles les seules causes de variation ?

Le nombre de tués sur la route (3)

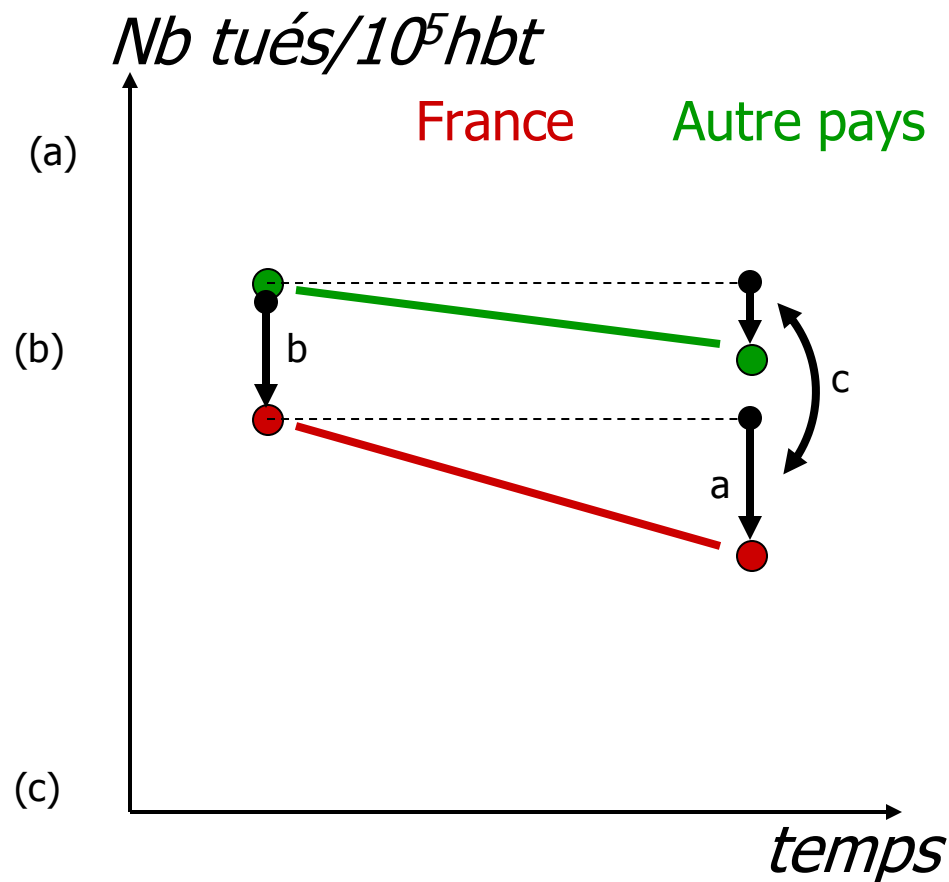


- Et maintenant comment le calcule-t-on :
 - En population totale ou par échantillonnage ?
 - Par recueil
 - prospectif (ad hoc)
 - rétrospectif (données de routine)
 - rétrospectif (enquête spécifique)
 - Par sources multiples ou unique ?
 - Peut-on identifier un goulet d'étranglement de l'information ?
 - Certitude sur la définition du « tué sur la route » ?

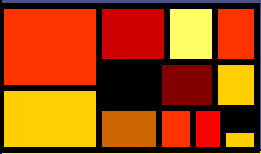
Le nombre de tués sur la route (4)



- Comparaison longitudinale : effet des politiques publiques
- Comparaison transversale : se positionner par rapport aux voisins
 - ... avec ses limites :
 - Il y a plus de tués en France qu'en Suisse
 - Il y a plus de tués par habitant en France qu'en Crête
- Comparaison longitudino-transversale



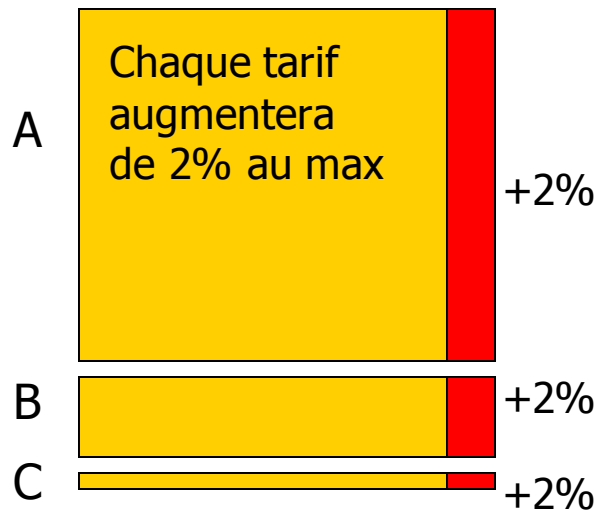
Evolution des tarifs autoroutiers



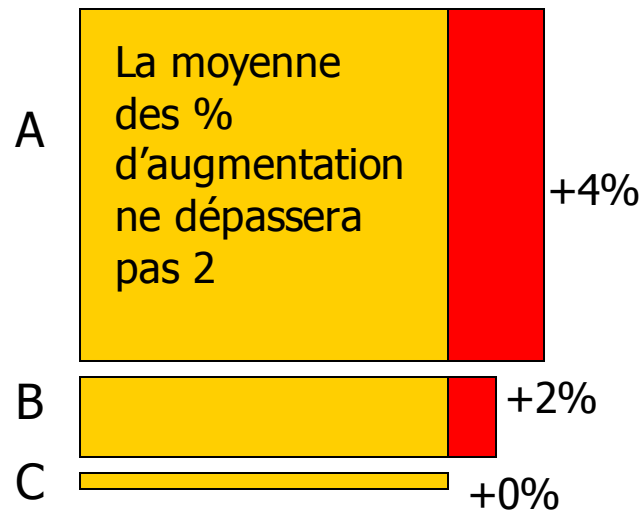
- « les tarifs des sociétés d'autoroutes augmenteront d'au maximum 2% par an »

Trois lectures plus ou moins honnêtes d'un objectif très flou :

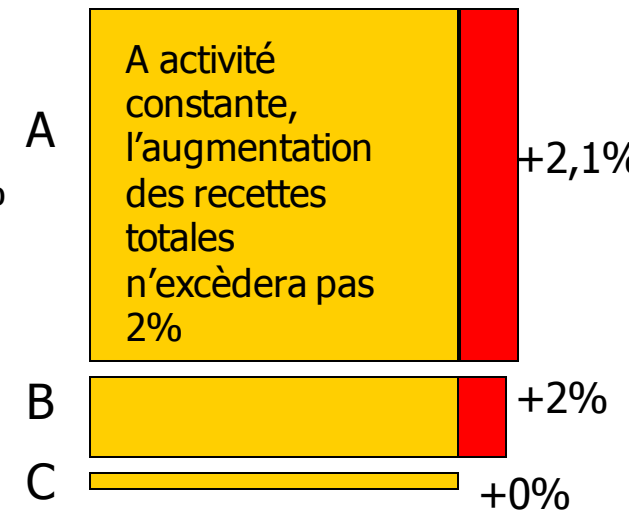
Lecture « chaque »



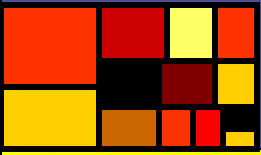
Moyenne faciale



Moyenne pondérée=CA



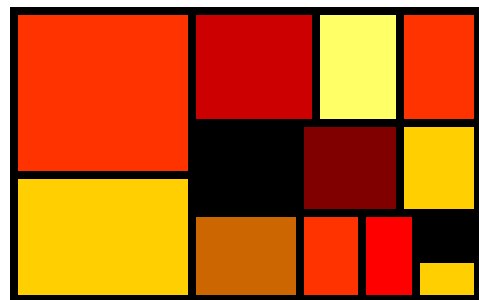
Statistiques trompeuses...



- Taux de redoublements dans des lycées :
 - « Dans le lycée d'en face, 50% des élèves redoublent leur terminale »
 - « Alors que notre lycée, en terminale, seulement 1 élève sur 3 est un redoublant »
- La délinquance stagne...
 - « Dans la commune gérée par nos adversaires, la délinquance a baissé de 40% la première année, puis a de nouveau augmenté de 40% l'année suivante. Bref, rien n'a changé. »



Indicateurs qualité nationaux, IPAQSS



- I. Définition
- II. Les indicateurs IPAQSS
- III. Plate-forme PLATINES

Définition

- Définition :
 - Variable synthétique (*nombre*) permettant d'évaluer (*quantifier*) l'adhésion d'un système ou d'une organisation à une norme (*valeur cible*)
- Type : souvent quantitatif
 - Proportion de dossiers conformes
 - Moyenne d'un score par dossier
- Interprétation en terme de « bien/mal », « pas assez/trop »...

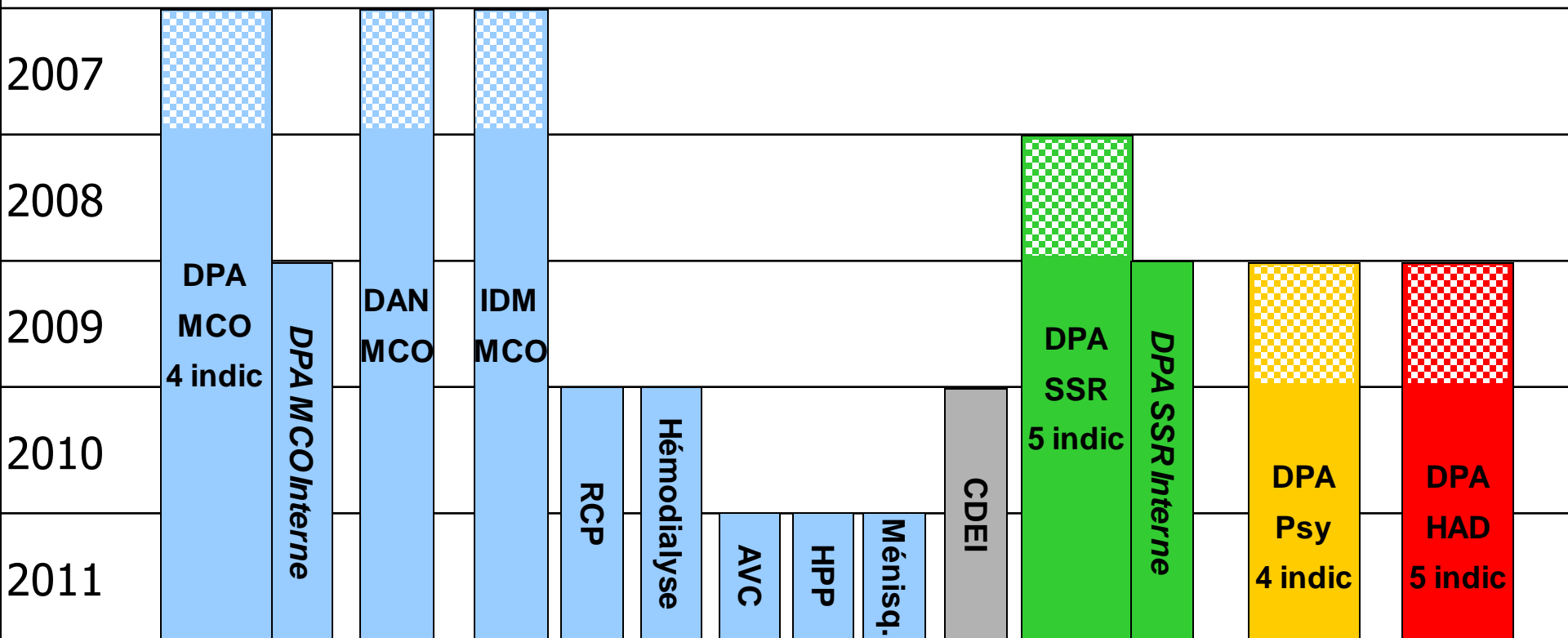
Exemples en Qualité

- Exemple d'indicateurs existants :
 - IPAQSS, DUQUE (qualité des soins)
 - SAFORA (satisfaction des patients)
 - ICALIN (infections nosocomiales)
- Exemples de normes ou référentiels
 - Recommandations de bonnes pratiques, recommandations de pratiques cliniques
 - Anciennement références médicales opposables
 - Code de Santé Publique
 - ex : envoyer les courriers de sortie dans les 8 jours
signer les prescriptions...*
 - COMPAQH (→ IPAQSS)

Les Indicateurs IPAQSS

- Définition
 - Indicateurs Pour l'Amélioration de la Qualité et de la Sécurité des Soins
 - Enquêtes nationales et obligatoires pour tous les établissements MCO (1144), SSR (1060), HAD, Psychiatrie
- Procédure :
 - Tirage au sort de dossiers patients
 - Examen des dossiers, notation
 - Calcul d'indicateurs de qualité
 - Amélioration itérative des indicateurs en collaboration avec les services
- Finalité = éléments de comparabilité :
 - Transversale : entre établissements à un instant donné
 - Longitudinale : suivi d'un établissement dans le temps
 - Intégration dans la certification v2010 (sept 2009)

Les indicateurs IPAQSS au fil du temps...



L'année correspond aux dossiers étudiés, et non à la date d'étude.

AVC : prise en charge des accidents vasculaires cérébraux

CDEI : conformité des demandes d'examen d'imagerie

DAN : dossier d'anesthésie

DPA : dossier patient

HPP : prévention et traitement des hémorragies du post partum

IDM : prise en charge hospitalière de l'infarctus du myocarde

IPAQSS : indicateurs pour l'amélioration de la qualité et de la sécurité des soins

RCP : traçabilité des réunions de concertation pluridisciplinaires

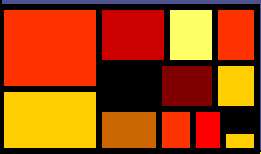
Exemple du délai d'envoi des courriers (DEC)

- Justifications :
 - courrier de fin d'hospitalisation = élément clé de la continuité des soins.
 - L'envoi des courriers de fin d'hospitalisation doit être effectué réglementairement dans un délai de 8 jours (Article R 1112-1 du CSP).
 - Il est aisé de mettre en place des actions pour améliorer les délais de rédaction et d'envoi des courriers
 - Cela devrait permettre une amélioration de la coordination entre les professionnels de santé pour le soin du patient
- **Exercice :** examinez le questionnaire et les résultats de l'indicateur n°1. Comment est-il calculé ?

Enseignements tirés du cas pratique

- Toujours se reposer sur une norme opposable
- Identifier le concept à contrôler
 - Ici, envoi rapide d'un courrier complet*
- Identifier l'entité de base : acte, séjour, patient ?
 - Ici on compte des séjours, ou plus précisément des sorties*
- Identifier les cas dans lesquels cette variable n'a pas de sens, i.e. les critères d'exclusion
- Identifier une variable composite permettant de contrôler ce concept : critères, méthode de calcul
- Exemple générique de méthode de calcul :
 - Dénominateur = nombre de cas où la règle est en théorie applicable ?
 - ex : séjours avec sortie au domicile*
 - Numérateur = parmi ces cas, nombre de cas qui satisfont le critère de qualité ?
 - Calcul = numérateur / dénominateur

Plateforme PLATINES



- Site web accessible au grand public :
<http://www.platines.sante.gouv.fr/>
- Présente pour le grand public
 - Des indicateurs d'activité
 - Les indicateurs IPAQSS
 - Les indicateurs ICALIN
- Exemples ci-après pour le CHRU de Lille



MINISTÈRE DE LA SANTÉ
ET DES SPORTS

DREES-DGOS

PLATINES : PLATeforme d'INformations sur les Etablissements de Santé



imprimer



accueil

- ▶ Comment naviguer sur le site ?
- ▶ Sources et résultats ?

Recherche d'établissements

Nom de l'établissement:

Nom de la commune:

Région:

tous



Département:

tous



Statut de l'établissement:

tous



Recherche avancée : 0 critère(s)



Chercher

Masquer les 1 résultat(s)

- ▶ CHU DE LILLE (59)

CHU DE LILLE

FINESS : 590780193

Centre hospitalier régional (CHR)

Exporter la fiche au format PDF Exporter les indicateurs qualité au format PDF

Etablissement public

LILLE (59)

Tous les indicateurs	Lits et Places	Equipements	Activité générale
Maternités	Urgences et réanimation	Autres activités autorisées	Autres prestations
Pathologies courantes	Chirurgie ambulatoire	Indicateurs Qualité	Environnement démographique

🔍 **Catégorie de référence:**

Centres Hospitaliers Régionaux

Centres Hospitaliers Régionaux

Lits et places (MAJ. : 8 février 2011)

Année 2009	Hospitalisation complète	Hospitalisation de jour
	🔍 Nombre de lits	🔍 Nombre de places
Médecine (M)	1021	187
Chirurgie (C)	809	68
Obstétrique (O)	143	35
Total MCO	1973	290
Psychiatrie	90	56
Soins de suite et réadaptation	209	15



Haut de page

Equipements (MAJ. : 8 février 2011)

Comparaison avec l'ensemble des CHR

Equipements (MAJ.: 8 février 2011)

Année 2009		Comparaison avec l'ensemble des CHR	
			% de oui
<input type="radio"/> Scanner (ou tomodensitomètre)	oui		97 %
<input type="radio"/> Imagerie par résonance magnétique (IRM)	oui		94 %
<input type="radio"/> Caméra à scintillation (ou gamma-caméra)	oui		81 %
<input type="radio"/> Tomographe à émission de positons (TEP) ou caméra TEP	oui		71 %
<input type="radio"/> Salles d'hémodynamique	non		81 %
<input type="radio"/> Salles de coronarographie	oui		97 %

[Pour comprendre](#)

[Haut de page](#)

Activité générale (MAJ.: 8 février 2011)

Année 2009		Hospitalisation		Comparaison avec l'ensemble des CHR	
	Nbre de séjours	complète Nbre de séjours	de jour Nbre de séjours		Valeur
Médecine (M)	50218	50218	35752		
Chirurgie (C)	29825	29825	9149		
Obstétrique (O)	6402	6402	3891		
Total MCO	86445	86445	48792		
<input type="radio"/> Durée moyenne de séjour (en jours)		6.1			6.5
<input type="radio"/> Indice de performance		1.0			1.0

[Pour comprendre](#)

[Haut de page](#)

Maternités (MAJ.: 8 février 2011)

Année 2009		Comparaison avec l'ensemble des CHR	
			Valeur
<input type="radio"/> Centre périnatal de proximité	non		
<input type="radio"/> Niveau de maternité			
<input type="radio"/> Maternité de niveau 1	non		
<input type="radio"/> Taux de césariennes (niveau 1)	- %		21.5%
<input type="radio"/> Maternité de niveau 2	non		
<input type="radio"/> Taux de césariennes (niveau 2)	- %		23.7%
<input type="radio"/> Maternité de niveau 3	oui		
<input type="radio"/> Taux de césariennes (niveau 3)	15.5 %		21.4%
Nombre annuel d'accouchements	5091		
<input type="radio"/> Durée moyenne de séjour des accouchements normaux (en jours)	4.9		4.2

[Haut de page](#)

Lutte contre les infections nosocomiales (MAJ. : 1er juillet 2010)

Année 2008	Résultat de l'établissement	Positionnement de l'établissement (Classe de performance)	Évolution par rapport à l'année précédente	Atteinte de l'objectif national (oui/non) sans objet
Le score agrégé (score sur 100)	92	A	→	sans objet
ICALIN (Indice composite des activités de lutte contre les infections nosocomiales) (score sur 100)	100	A	→	oui
ICSHA (Indicateur de consommation de produits hydro alcooliques) (score sur 100)	208.2 %	A	→	oui
SURVISO (Surveillance des infections du site opératoire) (réalisation d'une enquête d'incidence)	oui	2/10	→	oui
ICATB : Indice Composite de bon usage des AnTiBiotiques (score sur 20)	18	A	→	oui

[Pour comprendre](#)

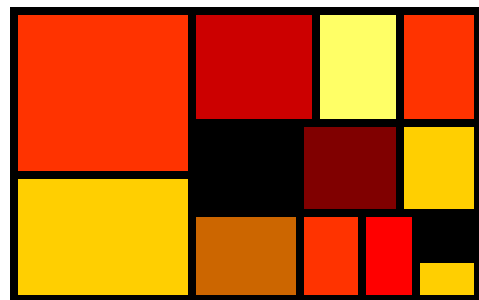
[Haut de page](#)

Qualité de la prise en charge (MAJ. : 1er juillet 2010) HAS

Année 2009	Résultat de l'établissement (valeur + fourchette)	Positionnement de l'établissement par rapport à la moyenne nationale	Évolution par rapport à l'année précédente	Atteinte de l'objectif national
Tenue du dossier patient (TDP) (score sur 100)	66 ±4	☹	→	NON
Délai d'envoi du courrier de fin d'hospitalisation (DEC) (pourcentage de dossiers conformes)	40% ±11%	☹	→	NON
Traçabilité de l'évaluation de la douleur (TRD) (pourcentage de dossiers conformes)	34% ±10%	☹	↗	NON
Dépistage des troubles nutritionnels (DTN) (pourcentage de dossiers conformes)	74% ±10%	☹	→	OUI
Tenue du dossier anesthésique (DAN) (score sur 100)	68 ±4	☹	→	

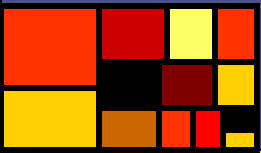
[Pour comprendre](#)

Dossier patient électronique



- I. Définition et contenu du dossier médicale
- II. Dossier informatisé et terminologies : contenu minimal lié au PMSI
- III. Contenus optionnels : médicaments, biologie

Définition du Dossier Médical

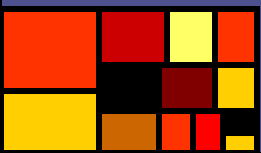


- En théorie
 - Ensemble des documents conservés par un médecin ou le service hospitalier qui le traite : données subjectives, données objectives, données interprétatives
 - Les notes personnelles, confidentielles qui appartiennent au seul médecin qui les a prises (famille, travail, vie sentimentale...)
- En pratique, deux organisations très différentes :
 - Soit **dossier patient unique** : 1 patient = 1 grosse enveloppe, conservée par l'établissement
 - Soit superposition de (souvent cliniques privées) :
 - un **dossier séjour** qui appartient à l'établissement (avec les résultats d'examens, comptes-rendus, courriers) → sans historique
 - plusieurs **dossiers patients**, un pour chaque médecin, avec essentiellement les notes personnelles et historique, mais pas de transdisciplinarité.

Utilité du dossier médical

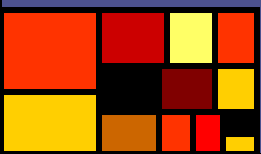
- Principalement assurer la continuité des soins
- Mais aussi
 - Évaluation de l'activité hospitalière
 - Quantitative (médico-économique)
 - Qualitative (accréditation, évaluation des pratiques, IPAQSS...)
 - Information épidémiologique et de santé publique
 - Preuve médico-légale
 - En cas de litige, le médecin devra apporter la preuve que l'information a été délivrée au patient. Cette preuve peut être apportée par tout moyen. (Art.L1111-2)

Contenu du dossier médical d'un patient hospitalisé (Art. R 710-2-2)



- Informations à l'admission
 - Lettre du médecin
 - Observation :
 - Motifs d'hospitalisation
 - Antécédents et facteurs de risque
 - Conclusions de l'évaluation clinique initiale
 - Prise en charge :
 - Soins et prescriptions à l'entrée
 - Soins et prescriptions lors de la consultation externe ou aux urgences

Contenu du dossier médical d'un patient hospitalisé (Art. R 710-2-2)



- Informations au cours du séjour
 - Observations cliniques
 - Prise en charge
 - Soins, prescriptions
 - Examens complémentaires
 - Dossiers spécifiques
 - Dossier d'anesthésie
 - Compte-rendu opératoire ou d'accouchement
 - Consentement écrit du patient si requis
 - Actes transfusionnels et fiche d'incident transfusionnel
 - Dossier de soins infirmiers
 - Correspondances entre professionnels de santé

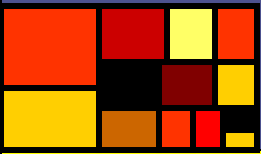
Contenu du dossier médical d'un patient hospitalisé (Art. R 710-2-2)

- Informations en fin de séjour
 - Compte rendu d'hospitalisation et lettre de sortie (adressée au médecin traitant)
 - Prescription de sortie et les doubles d'ordonnance de sortie
 - Modalités de sortie (domicile, transfert...)
 - La fiche de liaison infirmière
- Obligations générales
 - Pour toute pièce du dossier
 - Date
 - Identité du patient (nom, prénom, date de naissance ou numéro d'identification ou étiquette)
 - Identité du professionnel de santé qui a recueilli ou produit les informations
 - Prescriptions médicales
 - Date et heure
 - Identité lisible du médecin et signature

Informatisation

- Le dossier prend des formes diverses
 - Papier manuscrit ou imprimé, films radiologiques...
 - informatisation partielle ou totale, DVD...
- Enjeux de l'informatisation :
 - Apport des données : accélération, centralisation, intégrité, délocalisation
 - Accès aux données : accélération, multiplication (disponibilité) et délocalisation
 - Exploitation : automatisation ou au moins accélération
 - Stockage : diminution du coût, de l'altération avec le temps l'usage ou les accidents, disparition des transports

Recueil obligatoire dans le cadre des PMSI (MCO, SSR, HAD, Psy)



MCO

SSR

HAD

PSY

Administratif

Id, mvts

Id, mvts

Id, mvts

Id, mvts

Diagnostics

CIM10 :
DP, DR, DAS

CIM10 :
FP, MMP, AE, DAS

MP, MA, MD
CIM10 : DP, DAS

CIM10 : DP, DA

Actes méd.

CCAM

CCAM

CCAM

-

Actes autres

-

Activités de rééducation-réadaptation

-

Modalités

Consommables

DMI, MON

MON

-

-

Dépendance

-

AVQ

AVQ & Karnofsky

AVQ

Groupage MCO

Groupage SSR

Groupage HAD

séjour en GHS
=> € / séjour

journées en GMD
=> bientôt € / jour

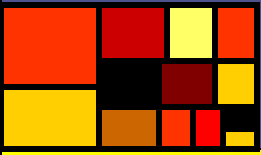
journées en GHPC
=> € / jour

Pas de groupage

Terminologies

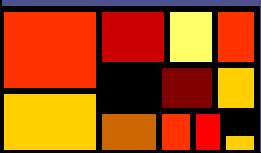
- Permettent de remplacer le vocabulaire :
 - Associent des codes et des libellés
 - Permettent, pour un concept donné, de trouver le code correspondant
 - Contrairement au texte libre, les codes sont sans ambiguïté
- → Support de l'interopérabilité sémantique
- Exemple :
 - Érysipèle = érésipèle = dermo-hypodermite = A46 en terminologie CIM10

Les terminologies utilisées en PMSI



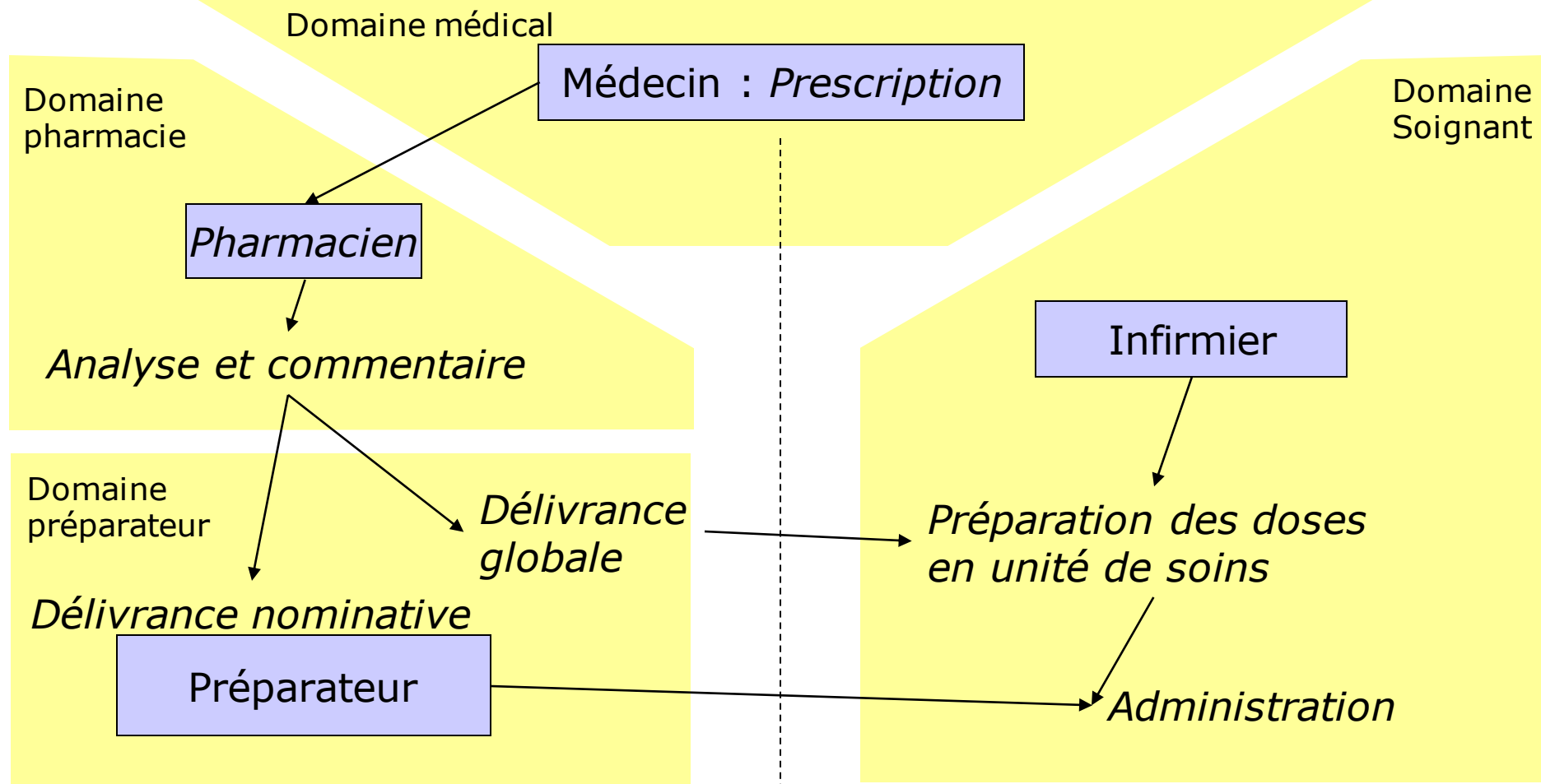
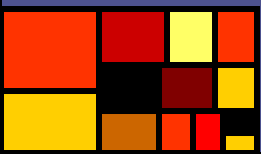
- Codage des diagnostics médicaux :
 - CIM10 en MCO, HAD, SSR, PSY (et codes spécifiques HAD pour les modes de prise en charge)
 - (international : CIM9, SNOMED, MESH, ICTC... Psy : CFTMEA...)
- Codage des actes médicaux :
 - CCAM (MCO, SSR)
 - (anciennement : CdAM. International : CIM9, SNOMED-CT...)
- Codage des actes de rééducation et réadaptation (SSR)
- Codage de la dépendance :
 - AVQ (SSR, HAD, PSY), Indice de Karnofski (HAD)
 - mais aussi les grilles GCAP et AGIRR
- Consommables :
 - LPP pour les DMI
 - UCD pour les MON

Autres éléments facultatifs du dossier électronique

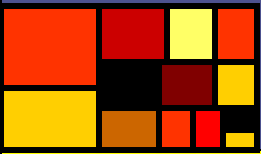


- Assez souvent :
 - Résultats de biologie
 - Courriers et comptes-rendus en texte libre
 - Applications métier hétérogènes : urgences, blocs opératoires, anesthésie, etc.
 - Serveur d'imagerie
- Plus rarement :
 - Médicaments prescrits, médicaments administrés
- Exceptionnellement :
 - Résultats structurés des systèmes de surveillance en réanimation (saturation, scope ECG...)
 - Enregistrements à la volée des DMI (exemple : défibrillateurs implantables...)

Circuit du médicament : prescription, délivrance/ dispensation, administration



Classifications de médicaments



- Nom commercial
 - Ex : Dafalgan
- DCI Dénomination commune internationale
 - Ex : Paracétamol / Acetaminophen
- Classifications diverses
 - dont la classification ATC
 - ATC = Anatomical Therapeutic and Chemical classification
 - Contient des codes, et une hiérarchie principalement basée sur l'indication thérapeutique

Exemple de l'aspirine dans la classification ATC

A alimentary tract and metabolism

- ↪ A01AD other agents for local treatment
 - ↪ A01AD05 ...aspirin...

B blood and blood forming organs

- ↪ B01AC platelet aggregation inhibitors
 - ↪ B01AC06 ...aspirin...

C cardiovascular system

- ↪ C10BX (...) other combinations
 - ↪ C10BX01 & C10BX02 ...aspirin...

M musculo-skeletal system

- ↪ M01BA anti-inflammatory
 - ↪ M01BA03 ...aspirin...

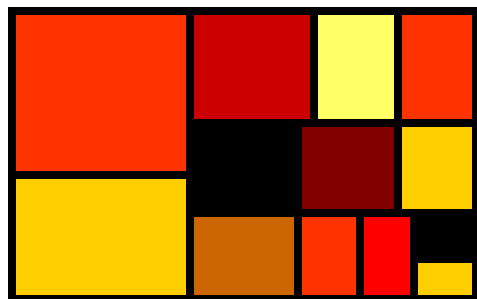
N nervous system

- ↪ N02BA salicylic acid and derivatives
 - ↪ N02BA01, N02BA51, N02BA71 ...aspirin...

Résultats de biologie

- Biologie : biochimie, hématologie, bactériologie, virologie, immunologie
- Exemples de terminologies :
 - LOINC (préférée par le ministère de la santé)
 - IUPAC
 - ...
- Mais la plupart du temps, chaque laboratoire produit sa terminologie maison :
 - Aucune interopérabilité sémantique :
 - Impossibilité de pooler deux hôpitaux
 - Dans un même hôpital, les données sont hétérogènes selon la date de mesure et parfois l'unité de laboratoire
 - Nécessité de produire des mappings *ad hoc*

Calculer des indicateurs dans le Dossier Patient électronique



- I. Pourquoi automatiser les indicateurs
- II. Méthodologie générale, problèmes rencontrés
- III. Tracer des événements dans le dossier patient électronique

Pourquoi construire des indicateurs automatisés depuis le dossier patient électronique ?

- Culture du papier dans les enquêtes qualités :
 - Souci d'un indicateur le plus proche possible du concept de qualité, avec forte interprétation humaine / experte
 - Recueil trop manuel, biaisé
 - Recueil fastidieux, source de tensions entre services
 - Biais d'échantillonnage flagrant
- Plaidoyer pour un calcul automatisé d'indicateurs qualité
 - Calcul à la volée, historique complet possible automatiquement (même rétrospectivement si l'indicateur change)
 - Automatique, sans effort marginal
 - Sans échantillonnage
 - Factuel, incontestable
 - Analyse plus poussée car données annexes disponibles pour réaliser des croisements statistiques

Méthodologie générale

Cas pratique : compléter cette démarche

1. Identifier un exemple de situation normale
« de qualité »
Exemple : toute hyperkaliémie > 5.5 doit être recontrôlée le lendemain au plus tard
2. En déduire les situations anormales
3. Décrire la méthode de calcul : numérateur, dénominateur
4. Fixer arbitrairement une valeur cible

Problèmes rencontrés

- Objectif : identifier des « mauvaises » raisons de déroger à la norme
- Application parfois décevante :
 - Identification de « bonnes » raisons de déroger à la norme
 - Si ces raisons sont justifiables alors que le score est mauvais, toute la démarche sera décrédibilisée !
- Attitude à avoir :
 - Humilité, présentation prudente des résultats
 - S'appuyer dessus comme un « outil de dialogue » tant qu'on est pas certain de la solidité
 - Test et revue des cas déviants, débogage de l'indicateur

Tracer des événements dans le dossier patient électronique

Cas pratique

Dans l'exemple suivant, sur quels éléments pouvez-vous vous appuyer pour détecter une hémorragie chez Mme Dupont ?

Dossier administratif

Femme de 88 ans

Diagnostics

- I10 Hypertension artérielle
- Z8671 Antécédents personnels de cardiopathies ischémiques
- I620 Hémorragie sous-durale non traumatique

Actes

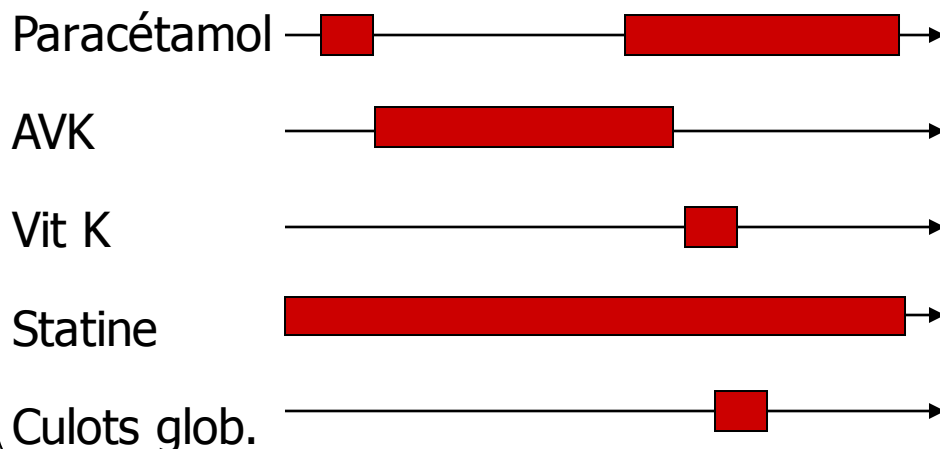
- ABJA002 Évacuation d'un hématome subdural aigu, par craniotomie
- FELF001 Transfusion de concentré de globules rouges

Courriers

Lettre de sortie

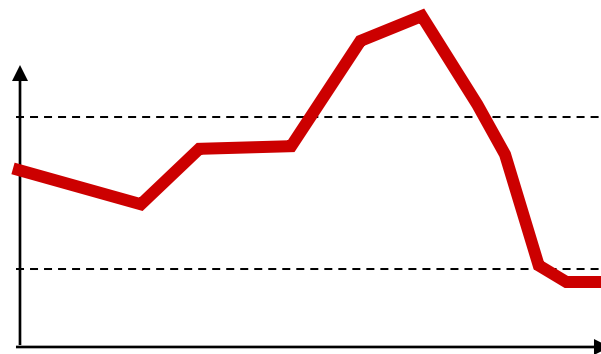
Compte-rendu opératoire

Médicaments

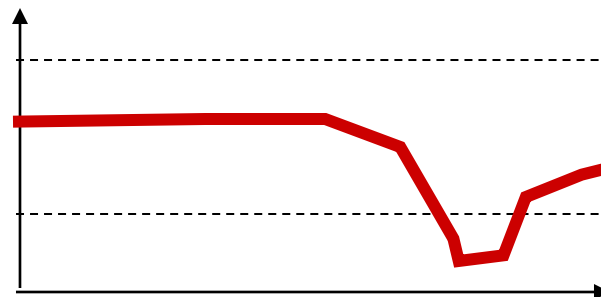


Biologie

INR



Hémo-globine



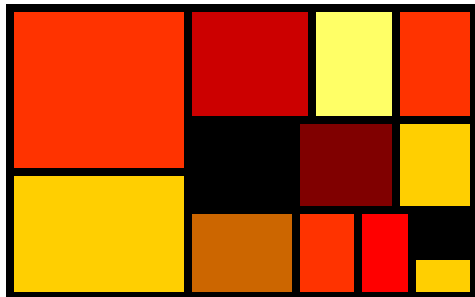
Tracer des événements dans le dossier patient



Dans l'exemple précédent, sur quels éléments pouvez-vous appuyer pour détecter une hémorragie chez Mme Dupont ?

- Dossier administratif :
- Diagnostics :
- Actes :
- Médicaments
 - Administrés : causes conséquences
 - Suspendus : causes conséquences
- Biologie : causes conséquences
- Courriers

Débogage d'indicateurs par Arbres de Décision



- I. Rappel sur notre indicateur
- II. Les arbres de décision
- III. Mise en pratique

Exemple d'indicateur

- Indicateur
 - Règle : une hyperkaliémie doit être recontrôlée le lendemain
 - Dénominateur : séjours avec hyperkaliémie
 - Numérateur : séjours avec hyperkaliémie et recontrôle le lendemain
 - Cas déviants = dénominateur \cap ($\overline{\text{numérateur}}$) :
parmi les hyperkaliémies, celles sans recontrôle le lendemain
- Revue des cas déviants
 - But = décrouvrir les facteurs qui, parmi les séjours contrôlés (dénominateur) prédisent la déviance (non appartenance au numérateur)
 - « bonne raison » de dévier : le patient est transféré entre-temps, ou dialysé
→ *il faut ajuster l'indicateur (restreindre le dénominateur)*
 - « mauvaise raison » de dévier : l'hyperkaliémie survient un dimanche
→ *l'indicateur est alors justifié*

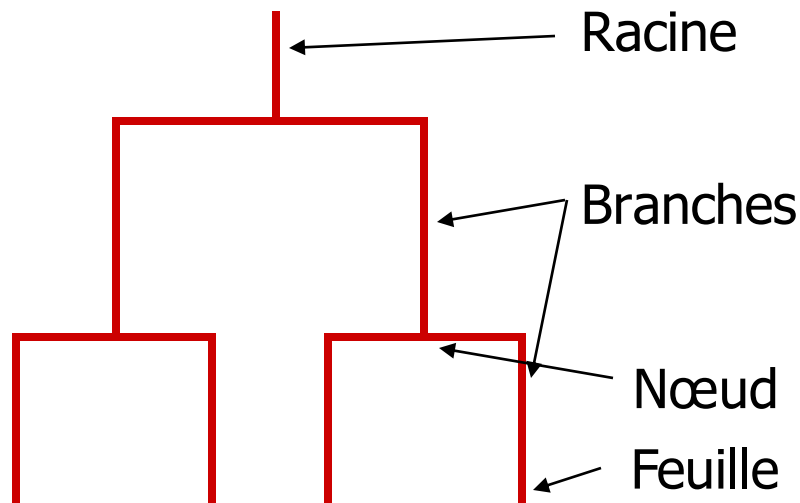
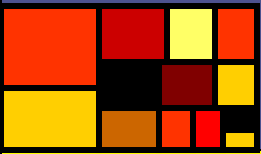
Data Mining (rappel)

- Définition
 - =KDD (Knowledge discovery in databases)
 - Mise en évidence de connaissances jusqu'alors inconnues dans des bases de données de grand dimension, à l'aide de méthodes dérivées de statistiques, du data management et de l'intelligence artificielle
 - Grande dimension :
 - Beaucoup de lignes (patients)
 - Beaucoup de colonnes (variables)
 - Beaucoup de cardinalité (tables et relations)
- Deux types :
 - Data mining non supervisé : motifs, associations
 - Data mining supervisé : expliquer une variable par les autres, prédire la valeur de cette variable

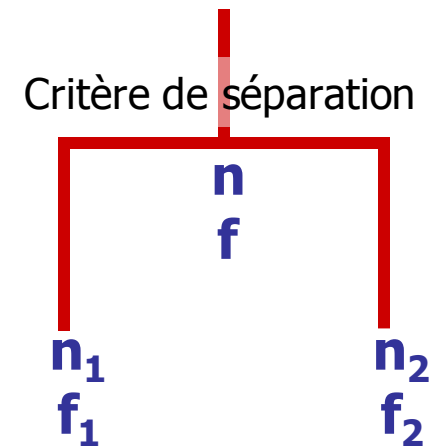
Les arbres de décision

- Méthode de data mining supervisé
- Variables explicatives : mélange de variables quantitatives, binaires, qualitatives
- Différents types selon la variable à expliquer :
 - Binaire : arbres de classification (les plus connus)
 - Qualitative : arbres de classification
 - Quantitative : arbres de régression
 - Quantitative (nombre d'événements) : arbres de Poisson
 - Censurée : arbres de survie

Principes des arbres binaires



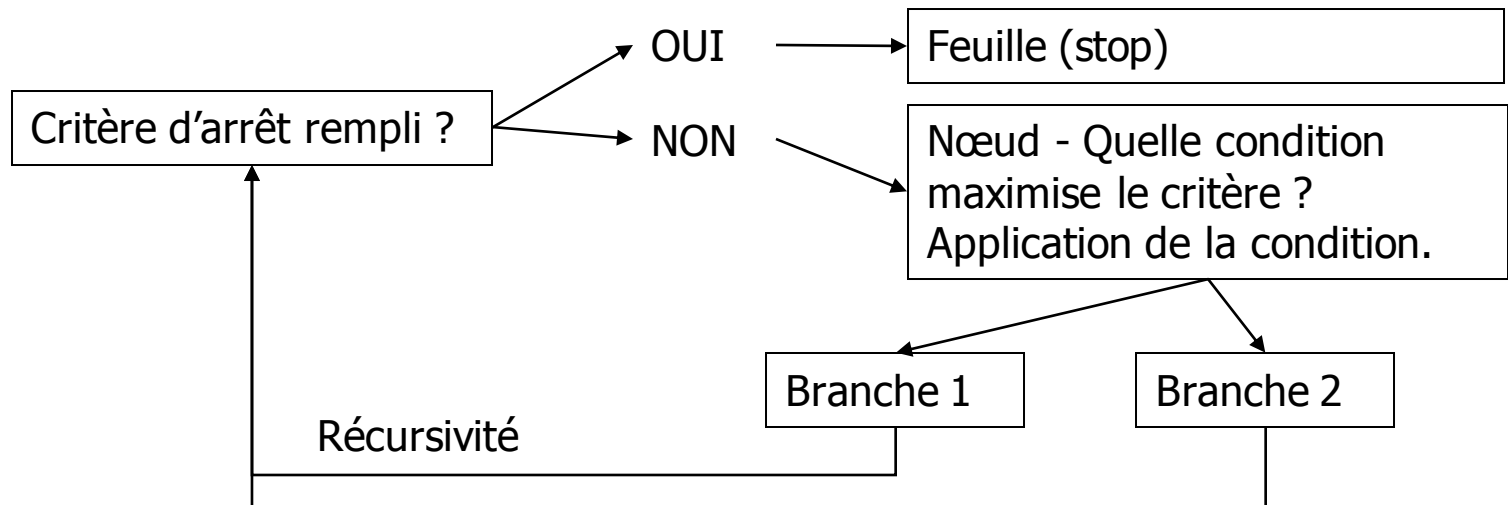
Unité de base : le nœud sépare un groupe en deux groupes, un avec une f augmentée, l'autre avec f diminuée.



$$n = n_1 + n_2$$
$$f = (n_1/n) \cdot f_1 + (n_2/n) \cdot f_2$$
$$\text{But : } f_1 \ll f \ll f_2$$

Principes des arbres binaires

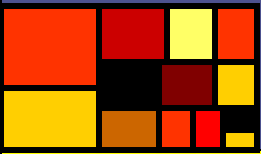
- Il s'agit donc d'un algorithme récursif : le produit d'une fonction est à son tour soumis à cette fonction
- Préalable :
 - Définir un critère de choix de la meilleure condition pour les nœuds
 - Définir un critère d'arrêt : si rempli, la branche devient une feuille. Sinon, l'algorithme continue.
- Déroulement récursif :



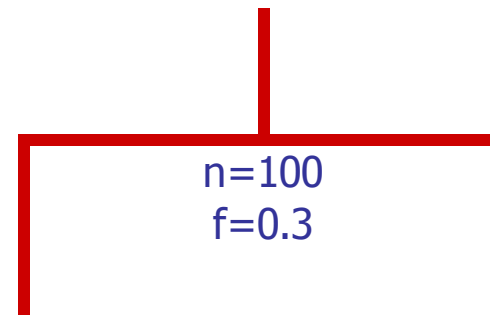
Approche par l'exemple

- Echantillon
 - tous les séjours avec une hyperkaliémie
 - $N=100$
- Variable à expliquer Y :
 - $Y=1$ si séjour déviant (pas de mesure le lendemain)
 - $Y=0$ si séjour normal (avec mesure le lendemain)
 - $f=P(Y)=0.3$
- Variables à expliquer X_i :
 - $X_1 =$ patient transféré $\{0;1\}$
 - $X_2 =$ l'hyperkaliémie est mesurée un dimanche $\{0;1\}$
 - $X_3 =$ patient insuffisant rénal $\{0;1\}$
 - $X_4 =$ patient diabétique $\{0;1\}$
 - ...
 - $X_{987} (...)$ $\{0;1\}$

Approche par l'exemple



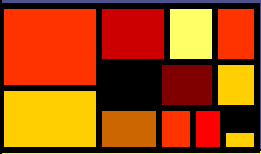
Première itération : nous cherchons la condition qui séparera le mieux les $Y=1$ (à droite) et les $Y=0$ (à gauche)



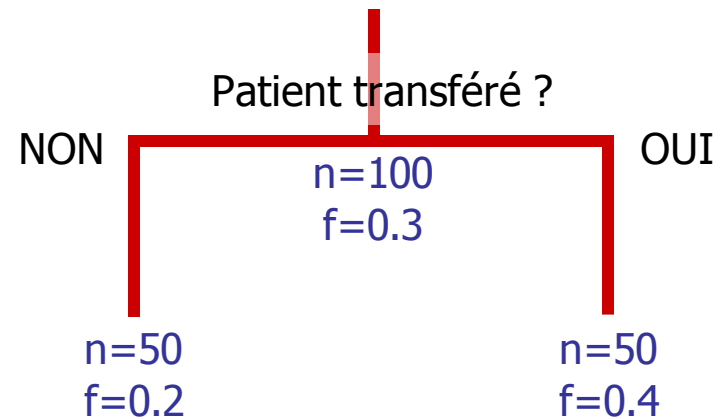
- Echantillon : $n=100$, $f=0.3$
- Variable à expliquer Y : séjour déviant ou non $\{0;1\}$
- Variables à expliquer X_i :
 - X_1 = patient transféré $\{0;1\}$
 - X_2 = dimanche $\{0;1\}$
 - X_3 = insuffisant rénal $\{0;1\}$
 - X_4 = diabétique $\{0;1\}$
 - ...

Variable	P du Chi ² vs Y
X1	1E-04
X3	1E-01
X2	0.15
X4	0.5
...	...

Approche par l'exemple

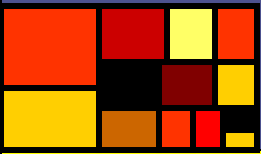


On choisit X_1 , qui est la plus fortement associée à Y .

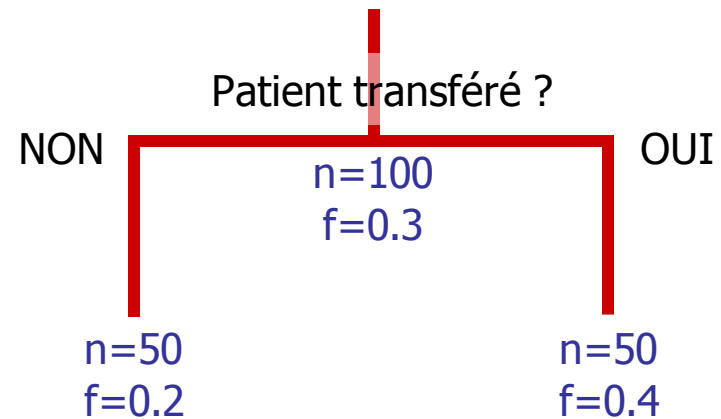


- Echantillon : $n=100$, $f=0.3$
- Variable à expliquer Y : séjour déviant ou non $\{0;1\}$
- Variables à expliquer X_i :
 - X_1 = patient transféré $\{0;1\}$
 - X_2 = dimanche $\{0;1\}$
 - X_3 = insuffisant rénal $\{0;1\}$
 - X_4 = diabétique $\{0;1\}$
 - ...

Approche par l'exemple



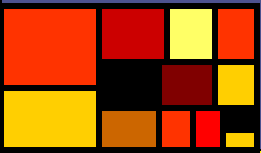
Itération maintenant sur la branche de droite.



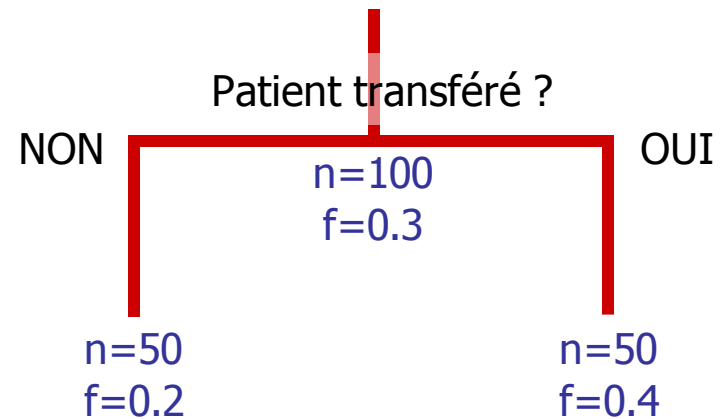
- Echantillon : $n=100$, $f=0.3$
- Variable à expliquer Y : séjour déviant ou non $\{0;1\}$
- Variables à expliquer X_i :
 - X_1 = patient transféré $\{0;1\}$
 - X_2 = dimanche $\{0;1\}$
 - X_3 = insuffisant rénal $\{0;1\}$
 - X_4 = diabétique $\{0;1\}$
 - ...

Variable	P du Chi ² vs Y
X3	0.21
X2	0.35
X4	0.60
...	...

Approche par l'exemple



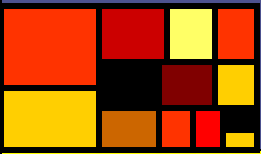
Itération maintenant sur
la branche de gauche



- Echantillon : $n=100$, $f=0.3$
- Variable à expliquer Y : séjour déviant ou non $\{0;1\}$
- Variables à expliquer X_i :
 - X_1 = patient transféré $\{0;1\}$
 - X_2 = dimanche $\{0;1\}$
 - X_3 = insuffisant rénal $\{0;1\}$
 - X_4 = diabétique $\{0;1\}$
 - ...

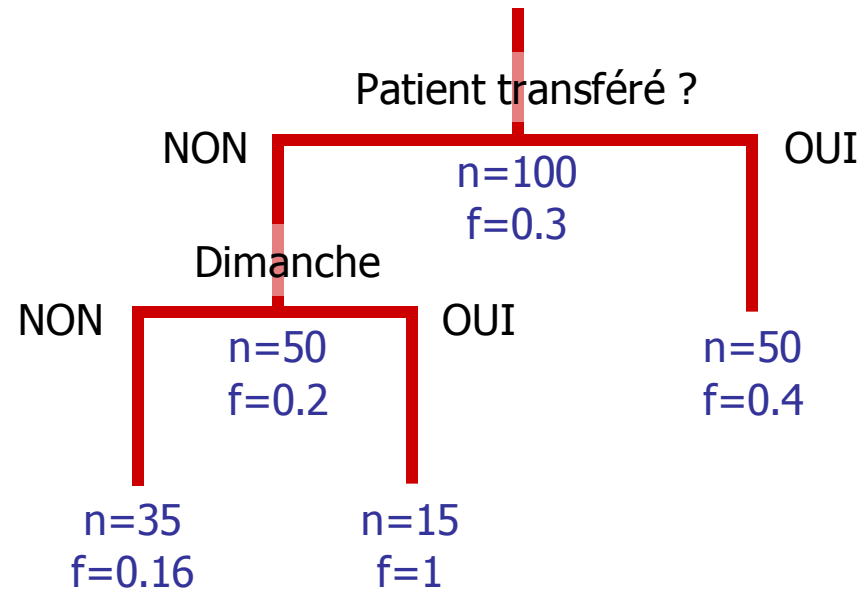
Variable	P du Chi ² vs Y
X2	1E-02
X3	0.36
X4	0.70
...	...

Approche par l'exemple

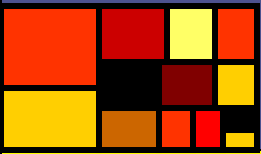


On choisit X2.
Peut-on encore continuer
?

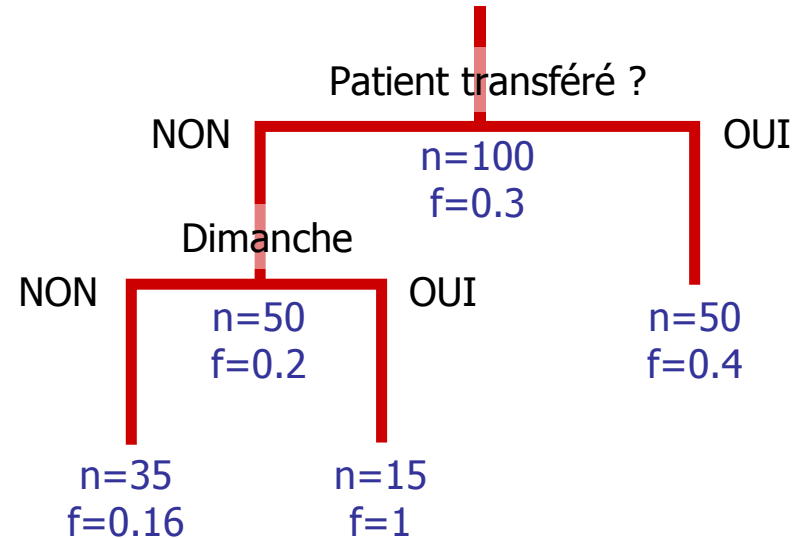
- Echantillon : $n=100$, $f=0.3$
- Variable à expliquer Y : séjour déviant ou non $\{0;1\}$
- Variables à expliquer X_i :
 - X_1 = patient transféré $\{0;1\}$
 - X_2 = dimanche $\{0;1\}$
 - X_3 = insuffisant rénal $\{0;1\}$
 - X_4 = diabétique $\{0;1\}$
 - ...



Approche par l'exemple



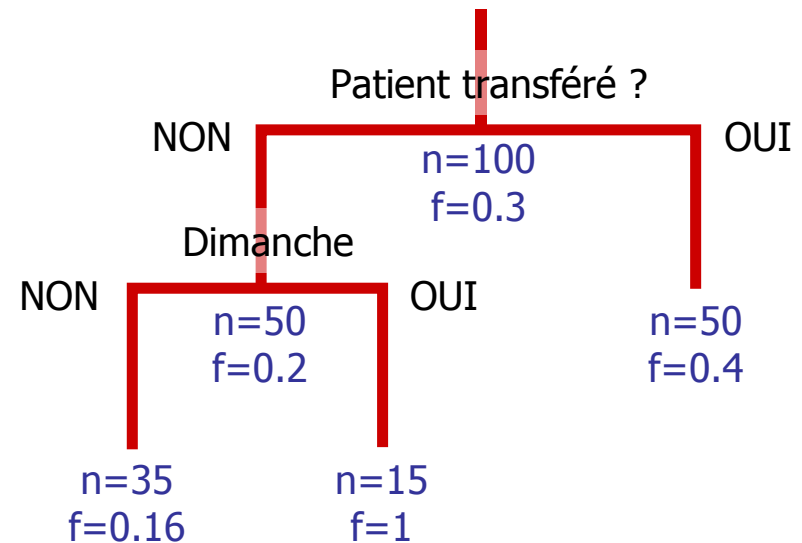
Arbre à 3 feuilles => 3 règles
de classification.
De droite à gauche :



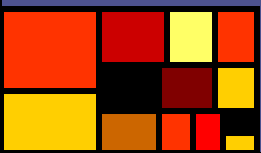
- Racine (tous les séjours) : $p(Y=1)=0.3$
- Patient transféré $\rightarrow p(Y=1)=0.4$
- Patient non transféré ET mesure le dimanche $\rightarrow p(Y=1)=1$
- Patient non transféré ET mesure autres jours $\rightarrow p(Y=1)=0.16$

Approche par l'exemple

- Retour au cas pratique
- Dans ce cas, quelle utilisation peut-on faire des connaissances mises en évidence ?

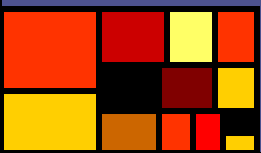


Utilisation des règles issues d'un arbre ou de règles d'associations



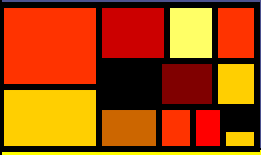
- Sous-groupes identifiés par des règles :
 - Chaque règle est un set de conditions menant à un effet $C_1 \cap C_2 \cap \dots \cap C_k \rightarrow E$
 - Soit C le set de conditions : $C = C_1 \cap C_2 \cap \dots \cap C_k$
 - Effet E : ici effet binaire $Y=1$
 - Condition C_i construite de différents manières :
 - Soit avec X_i binaire : par exemple X_i
 - Soit avec X_i quantitatif : par exemple $X_i > 50$
 - Soit avec X_i qualitatif : par exemple $X_i \in \{A;B\}$
- Quantités calculables :
 - Prévalence (confiance à la racine) : $P(E)$
 - Confiance (prévalence dans une feuille) : $P(E|C)$
 - Support (importance des cas de la feuille dans l'échantillon de départ) : $P(E \cap C)$
 - Lift (dans la feuille, la prévalence est multipliée par...) : $P(E|C) / P(E)$
 - Risque relatif (idem mais par rapport aux individus hors de la feuille ; moins pertinent si plusieurs règles sont utilisables) : $P(E|C) / P(E| \bar{C})$

Quelles questions nous sommes-nous posées pendant la construction ?



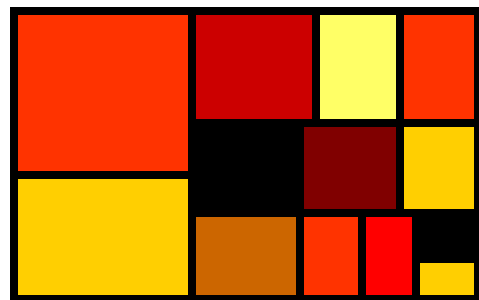
	Méthode CHAID	Méthode CART
Nature de Y	Binaire, qualitative	Binaire, quantitative, survie, Poisson
Nature de X	Binaire, qualitative, quantitative	Binaire, qualitative, quantitative
Critère d'arrêt	<ul style="list-style-type: none"> ■ Aucun χ^2 significatif ■ Y constant dans la branche ■ Toutes les X_i constantes dans la branche <p>=> Critères de pré-pruning</p>	<i>Post-pruning et cross-validation</i>
Critère de choix	Condition dont la variable maximise une quantité associée au χ^2	Indice de Gini

Avantages et inconvénients des arbres de décision



- Avantages :
 - Construction facile, paramétrable
 - Natures variées de Y
 - Natures variées de X, utilisables en même temps
 - Fonctionne en grande dimension (nombreuses lignes, nombreuses colonnes) sans difficulté
 - Arbre facile à interpréter, règles faciles à utiliser
- Inconvénients :
 - Risque de sur-ajustement (d'où la cross-validation)
 - Instabilité des arbres (mais prédiction peu altérée ; solution = *random forest trees*)
 - Moins performants que les réseaux de neurones
 - Résultats utilisables tels quels pour la prédiction, mais parfois peu montrables...

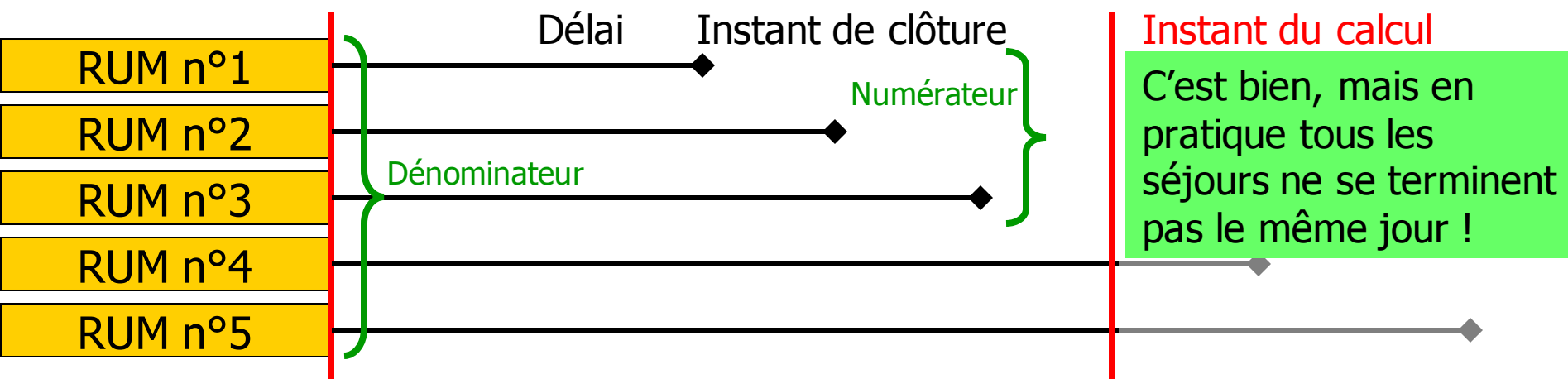
Faussement simple...



- I. Autour des délais et de l'exhaustivité (codage, émission de facture...)
- II. Autour des durées de séjour et des taux d'occupation
- III. Synthétiser l'origine géographique

Mesurer l'exhaustivité (de codage, d'émission...)

- Indicateurs numériques synthétiques
 - Exhaustivité de codage :
*proportion de RUM clôturés à une certaine date,
ou plutôt à un certain délai après la fin*

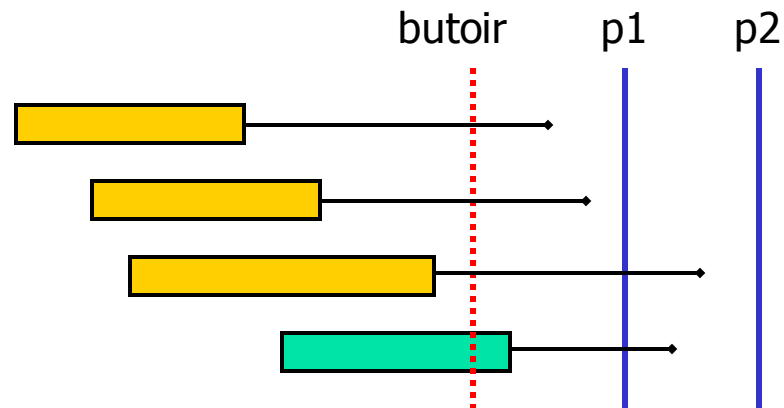


Mesurer l'exhaustivité de codage

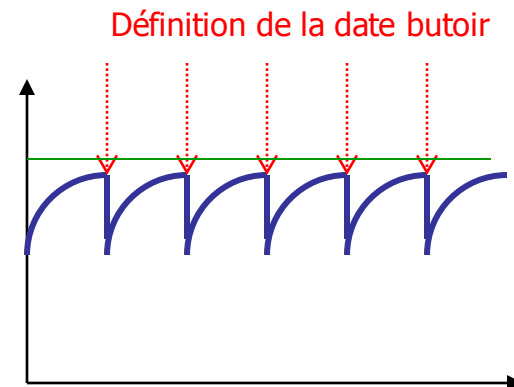
- Indicateurs numériques synthétiques

- Exhaustivité de codage

- Sur un groupe de séjour terminés avant une date butoir fixe
=> *chiffre croissant dans le temps, adapté à l'envoi par lots*
 - Assertions fortes : la date de référence est la date de sortie, on connaît le nombre de RUM exigibles



1- on définit une date butoir, tous les RUM terminés avant sont inclus. Les RUM terminés après, même codés, sont exclus.
2- à chaque date d'examen on calcule l'exhaustivité.
Résultats : 2/3 à p1, 3/3 à p2.

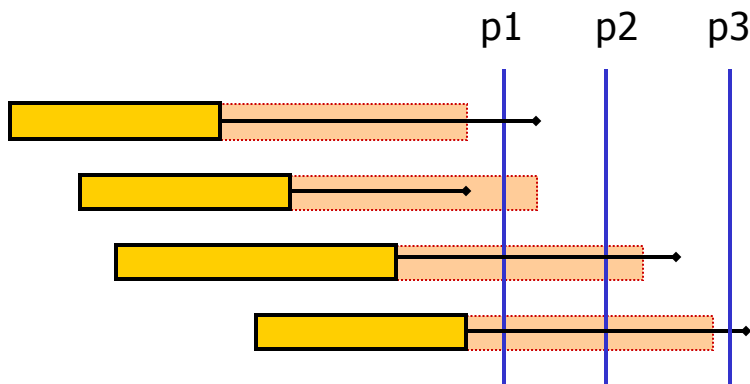


Mesurer l'exhaustivité de codage

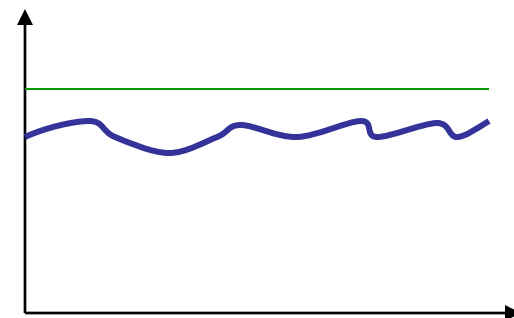
- Indicateurs numériques synthétiques

- Exhaustivité de codage

- Sur tous les séjours terminés au moins un délai avant la date d'observation
=> chiffre stable dans le temps, adapté à l'envoi au fil de l'eau
 - Assertions fortes : référence=sortie, nombre de RUM exigibles connu, existence d'un délai reconnu comme raisonnable (« le bien et le mal »)

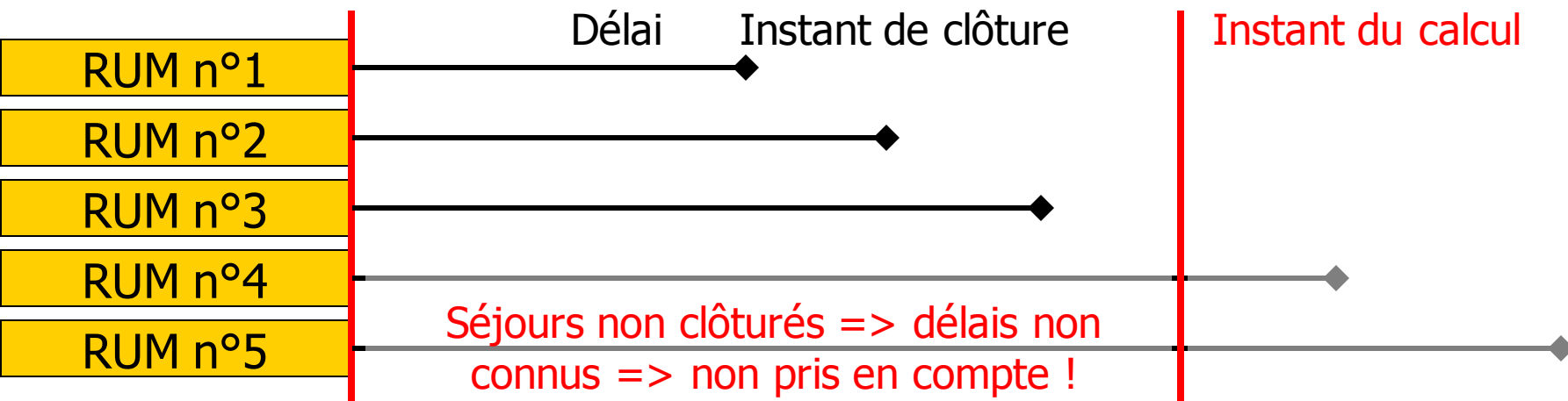


1- on définit un délai de référence
2- à chaque date de point on calcule l'exhaustivité en incluant seulement les RUM dont le délai est échu, même s'ils sont déjà codés.
Résultats: 0/1 à p1, 2/2 à p2, 3/4 à p3.



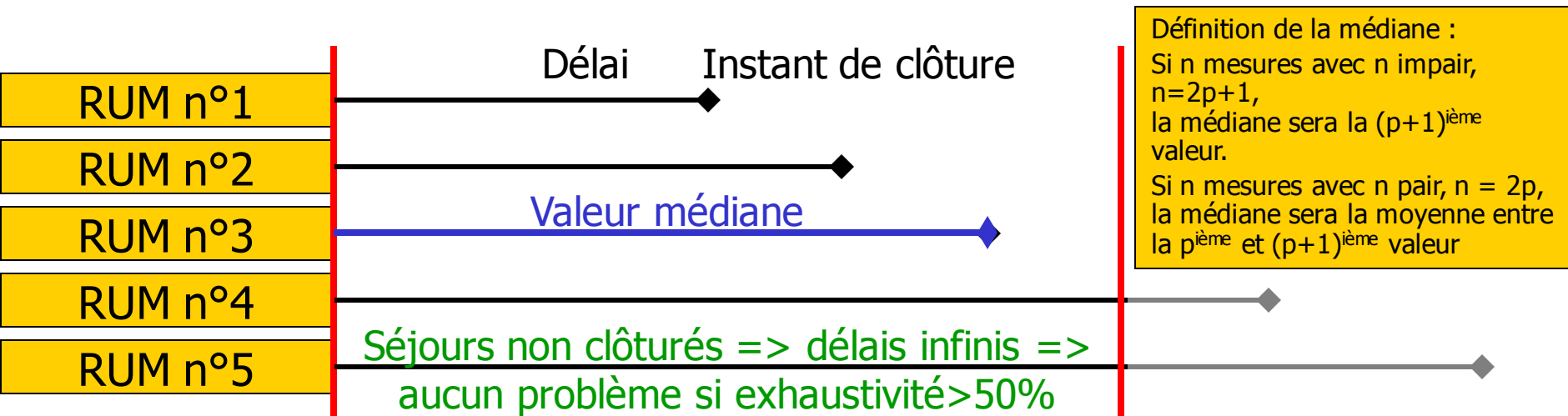
Mesurer l'exhaustivité de codage

- Indicateurs numériques synthétiques
 - Délai moyen de codage :
moyenne du délai entre la date de sortie et la date d'émission (clôture) d'un RUM



Mesurer l'exhaustivité de codage

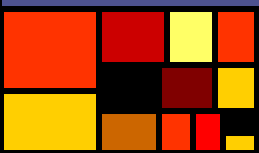
- Indicateurs numériques synthétiques
 - Délai médian de codage :
médiane du délai entre la date de sortie et la date d'émission (clôture) d'un RUM



Durée de séjour

- Calcul PMSI de la durée de séjour (parfois=0) :
 - Formule utilisée :
 - $DS = \text{date_sortie} - \text{date_entrée}$ (cf. représentation des dates par entiers)
 - Remarques
 - Attention : vaut zéro si le patient sort le jour même. Ce n'est donc pas une question de durée réelle, mais de position des E/S par rapport à minuit
 - Durée moyenne de séjour « $DMS = 0.2j$ » peut signifier par exemple « 80% des patients sortent le jour même et 20% sortent le lendemain ».
- Autres formules (toujours >0) :
 - Nombre de journées (assez pertinent)
 - Nombre de journées = $DS + 1$
 - Formule utilisée pour la valorisation des journées en réanimation
 - $\max(1, \text{date_sortie} - \text{date_entrée})$
 - Formule implicitement utilisée pour le calcul des tarifs en v11 (cf. post)
 - $\max(0.5, \text{date_sortie} - \text{date_entrée})$
- Dans une approche « recherche d'atypie », écart à la DMS (durée moyenne de séjour nationale dans le GHM concerné) :
 - $\text{Ecart_relatif} = \frac{[(DS+1)-(DMS+1)]}{[DMS+1]} = \frac{(DS-DMS)}{(DMS+1)}$

Performance globale : écart à la DMS en tenant compte des effectifs



$$\begin{aligned} &\text{Écart à la DMS ENC} \\ &= (DMS - DMS_{ENC}) / (DMS_{ENC} + 1) \end{aligned}$$

$$= \text{eff} * (DMS + 1)$$

$$= \text{eff} * (DMS_{ENC} + 1)$$

GHM	Effectif	DMS	DMS ENC	écart à la DMS ENC	nbg réels	nbg théo
01C01S	7	18.32	22.64	-18.3%	135.24	165.48
01C03V	13	14.75	11.96	21.5%	204.75	168.48
01C03W	13	28.75	26.05	10.0%	386.75	351.65
01C04V	65	16.66	12.86	27.4%	1147.9	900.9
01C04W	28	26.02	29.5	-11.4%	756.56	854
01C05V	17	10.76	10.7	0.5%	199.92	198.9
01C05W	5	14.61	16.64	-11.5%	78.05	88.2
01C06V	25	7.88	5.81	30.4%	222	170.25
01C06W	18	8.43	7.78	7.4%	169.74	158.04
01C08V	20	5.92	4.33	29.8%	138.4	106.6
01C08W	3	7.91	9.87	-18.0%	26.73	32.61
				-5.2%	3466.04	3195.11
					PF=	1.0847952

NB : la PG est robuste aux changements de classification des GHM

Interdiction de calculer la moyenne : effectifs inégaux

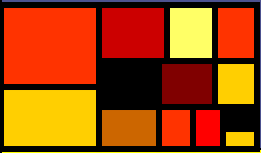
$$= \sum nbg / \sum nbg_{enc}$$

Calcul de l'occupation des UM

- En théorie
 - $\#journées_total = \sum (DS+1) = (\sum DS) + \#RUM$
 - $\#journées_max_théorique = \#lits_UM * 365$
 - $Taux\ d'occupation = \#journées_total / \#journées_max_théorique$
- En pratique méfiance...
 - Si un patient part le matin, un autre peut occuper le même lit dès midi. Il y a d'autant plus de chevauchements possibles que les passages sont brefs
 - Un taux d'occupation de 80% est jugé optimal en réanimation
 - Un taux d'occupation de 120% est habituel en chirurgie ambulatoire
 - L'interprétation du calcul tient aussi compte :
 - De l'organisation hebdomadaire du travail (fermeture le WE...)
 - De la gestion des vacances et jours fériés (...)
 - Le nombre de lits varie dans le temps (=> moyenne pondérée)

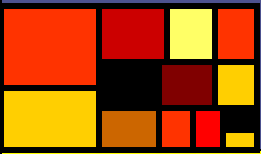
Notation :
#xxx = nombre de xxx

Représenter simplement le recrutement

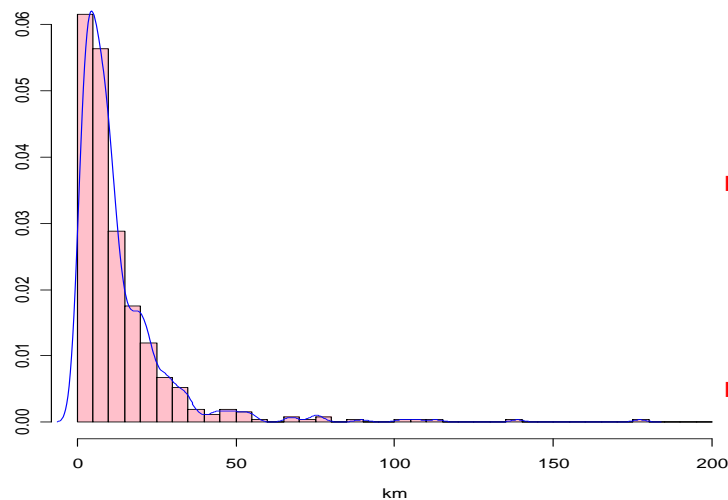


- Parts de marché :
 - Parmi les patients résidant sur un territoire défini
 - Sur une période de temps définie
 - Éventuellement pour une pathologie définie (ou plutôt un GHM ou regroupement de GHM)
 - $\frac{\text{Nb de séjours de ces patients produits par l'établissement}}{\text{Nb de séjours de ces patients produit par tout établissement français}}$
- Étendue du recrutement
 - Distance moyenne ou médiane entre l'établissement et le domicile du patient (cf. ci-après), par séjour
 - Plus robuste et facile à calculer : % de séjours pour lesquels le patient réside à l'extérieur d'un territoire de référence

Distance moyenne ou distance médiane ?



```
> km <- 8*exp(rnorm(n=500))
> hist(km, breaks=seq(from=0,to=200,by=5), cd="pink", freq=FALSE)
> lines(density(km), col="blue")
> summary(km)
  Min. 1st Qu. Median Mean 3rd Qu.  Max.
0.1485  3.7920  8.3400 14.0200 17.5700 150.8000
> mean(km>30)
[1] 0.104
> km <- c(km, 18000)
> summary(km)
  Min. 1st Qu. Median Mean 3rd Qu.  Max.
1.485e-01 3.792e+00 8.352e+00 4.992e+01 1.762e+01 1.800e+04
> mean(km>30)
[1] 0.1057884
```



- La moyenne n'est pas robuste :
 - La moyenne est trop sensible aux individus extrêmes
 - Or les distances ont habituellement des distributions lognormales
- Exemple sur données simulées
 - Soient 500 patients recrutés sur un territoire de proximité
moyenne=14 km, médiane=8.3 km, range=[0;150]
10.4% de distances>30km
 - Ajout d'un seul patient de Nouvelle Calédonie (en vacances chez sa sœur) : 18000km de l'hôpital
moyenne=50 km, médiane=8.3 km, range=[0;18000]
10.6% de distances>30km
 - Un seul patient sur 501 a le pouvoir de quadrupler la moyenne ! La médiane reste inchangée.