# Example 3: using artificial intelligence to detect and correct under-billing



Artificial intelligence, Data reuse, Big data in Healthcare

Pr Emmanuel Chazard



### Organization of the presentation

- Data quality
  - Missing values for multivalued encoded variables
- 2. Available data, data redundancy
  - Opportunities to identify missing diagnostic or procedure codes
- 3. From data quality to hospital incomes
  - Why adding diagnostic or procedure codes enable to earn money
- Description of the French PMSI files
  - ...Only if you intend to use those files
- 5. Quality control procedures
  - The most important part: expert or machine generated control rules, positive predictive values, metarules, etc.
- 6. Software for automated data quality control (2008)
  - Simple demonstration, and feedback



## Data quality failures (overview)



## Expectations about quality of non-missing data

- Each value is valid:
  - The value belongs to the right type
  - The value must be possible
  - For qualitative variables, the value must match a pattern or belong to a list of codes (terminology)
- The univariate distribution must be plausible
- The data must follow a correct bivariate (or conditional) distribution
  - Deterministic relationship (functional dependency)
  - Probabilistic relationship

<u>Counterexample :</u>						
Age="old" not numeric						
Age=834 out of range						
Diagnosis="HHFA001" more than 1 letter						
Diagnosis="B990" does not exist						
Each age belongs to [0;100] but mean=80 too high						
Each age belongs to [0;100] but SD=0 constant value						
Length of stay=2 not consistent Admission="2013-05-14" Discharge="2013-05-14"						
mean(age   unit="geriatrics")=21 too young						
mean(length of stay   admission mode="transfer")=1 too short						
correlation( age, length of stay) = -0.3 should be positive						

### What about missing values?

- Variables with 1 expected value: NAs are visible (at random or not...)
- Multivalued variables:
  - Often in healthcare, multivalued qualitative variables
  - E.g. diagnoses of Mr Dupont:
    - K35.9 appendicitis
    - I10 arterial hypertension
    - A given patient can have one or more diseases (separate table)
  - Our point of interest: the absence of a code
    - Is not directly detectable (no abnormal value, no "NA")
    - Can only be detected against a gold standard (e.g. free text discharge letter)
- Objective here: to identify and complete missing codes for multivalued variables



### Improving data quality may increase hospital incomes

#### Before correction

43 years old man Length of stay: 9 days

Principal diagnosis: N10 Acute pyelonephritis

#### After correction

43 years old man Length of stay: 9 days

Principal diagnosis:

N10 Acute pyelonephritis

Associated diagnosis:

**B24** AIDS (HIV)

Before: 11M041

Urinary or renal infection, age>17, level 1

1223 €

25M02B

After:

Diseases due to HIV, with only 1 infectious complication

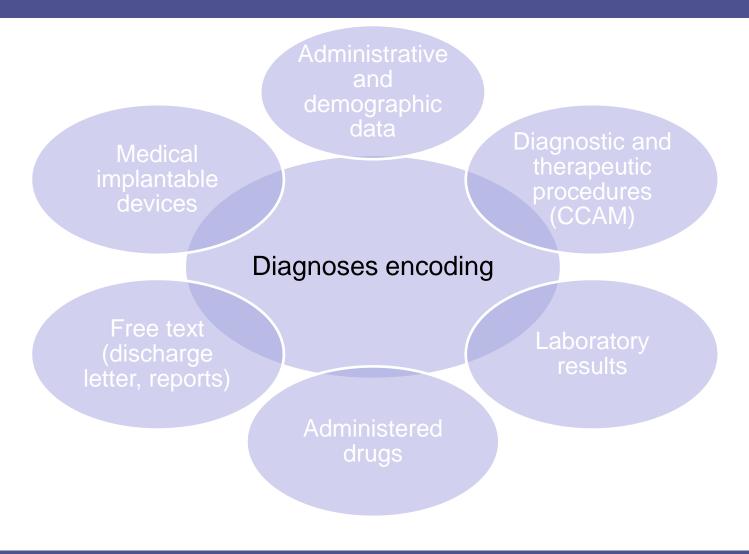
5135€

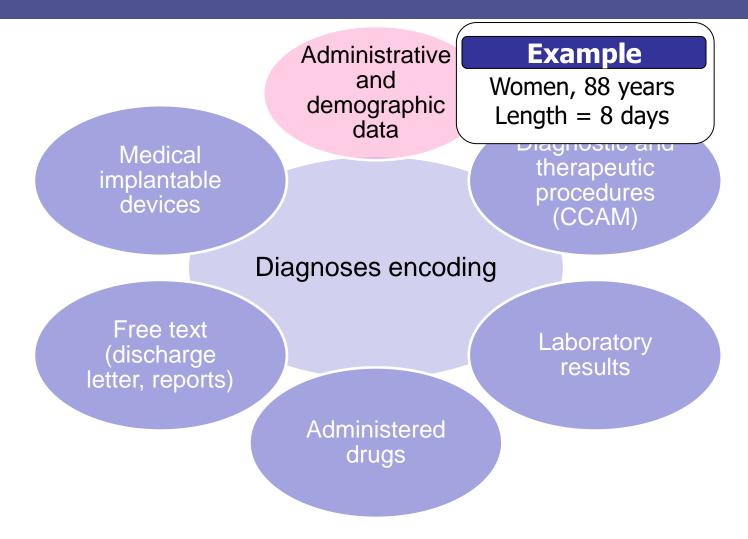
+3912€

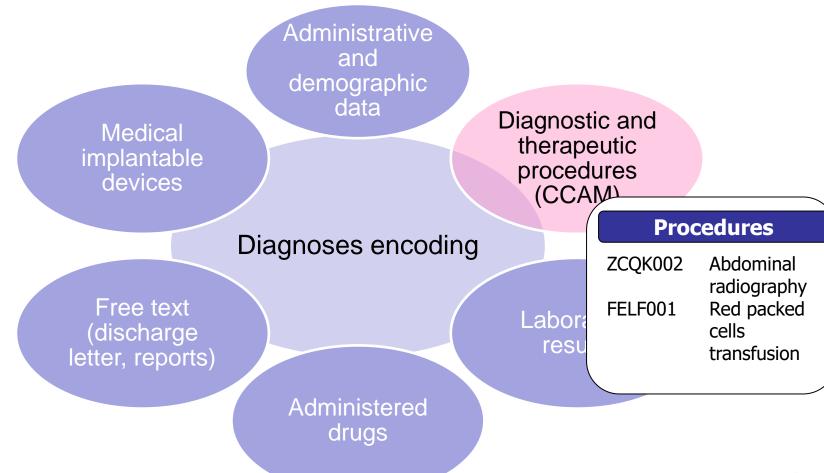


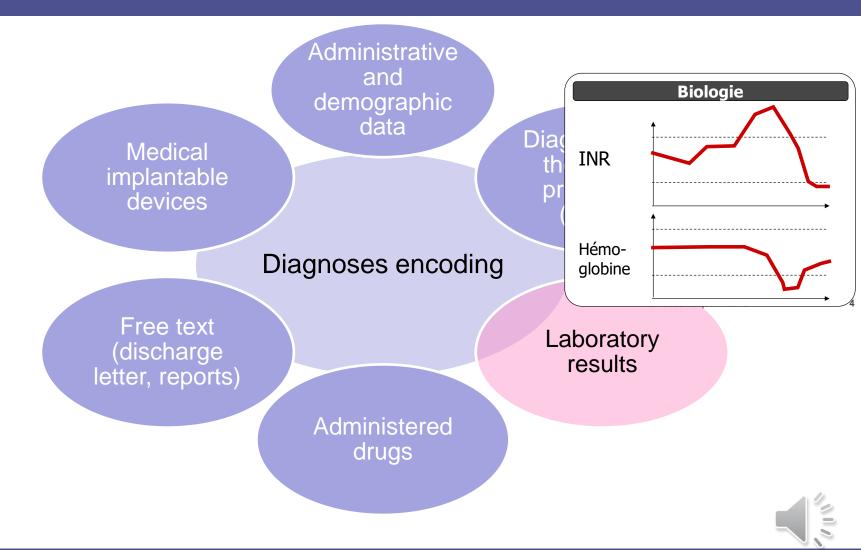
### Available data, Data redundancy

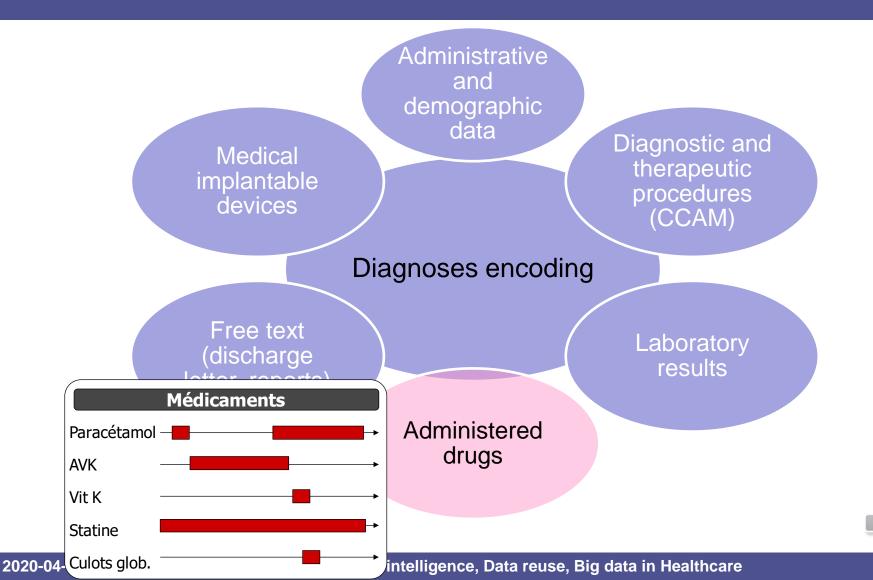


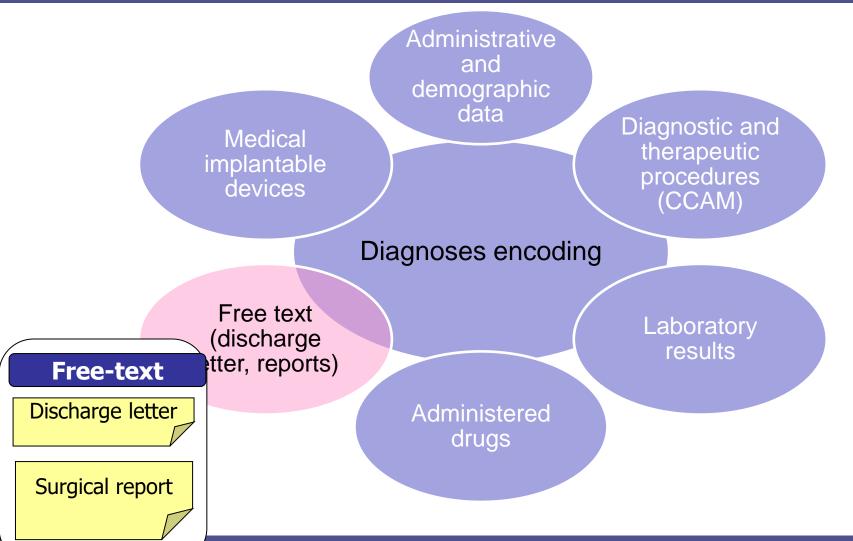














Medical implantable devices

Administrative and demographic data

Diagnostic and therapeutic procedures (CCAM)

Diagnoses encoding

Free text (discharge letter, reports)

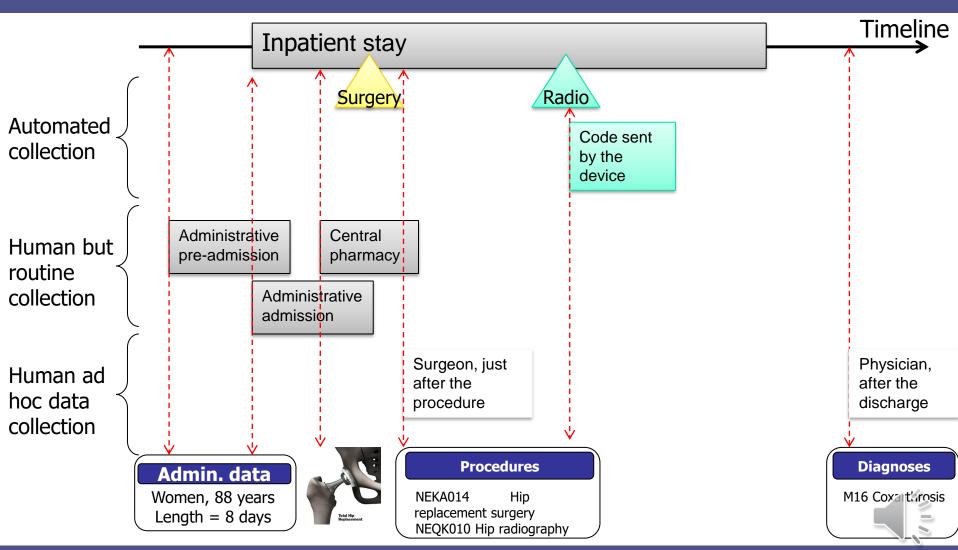
Administered drugs

Laboratory results



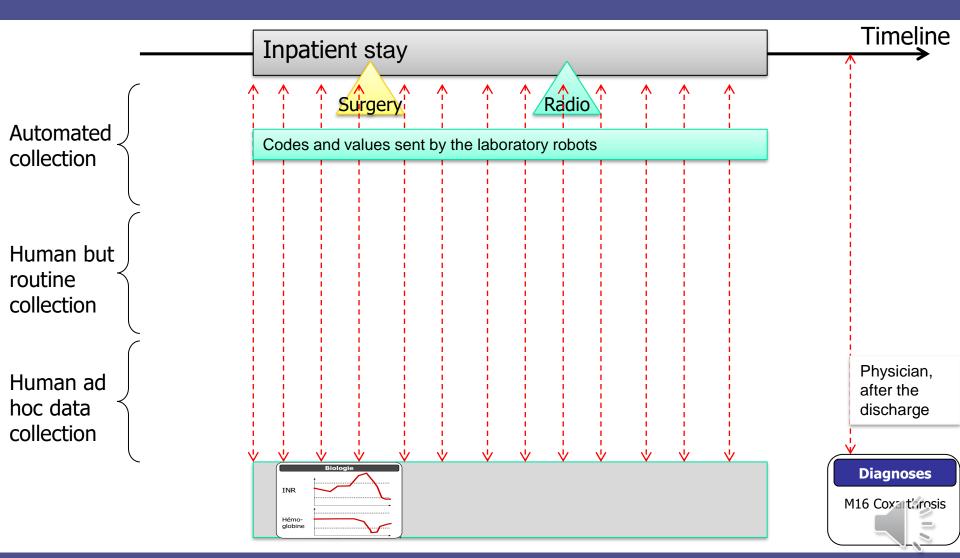
### Available information: temporal aspects

E.g.: scheduled hospitalization for surgical hip replacement



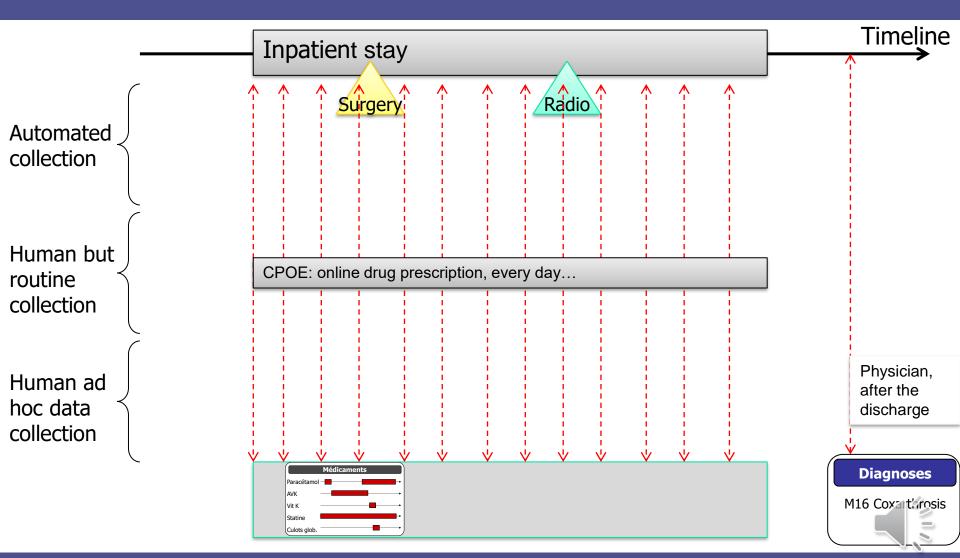
### Available information: temporal aspects

E.g.: scheduled hospitalization for surgical hip replacement



### Available information: temporal aspects

E.g.: scheduled hospitalization for surgical hip replacement



## Automated control of medical encoding: introduction

### Objective:

- Use every available data to correct the human ad hoc encoding
- Data to correct:
  - Diagnostic codes
  - Procedure codes
- Available data:
  - Administrative and demographic data
  - Drugs
  - Laboratory results
  - Procedure codes
  - Implantable medical devices
- Software developed for consultants
  - Uses artificial intelligence to automatically learn associations between those data types.



## From data quality to hospital incomes



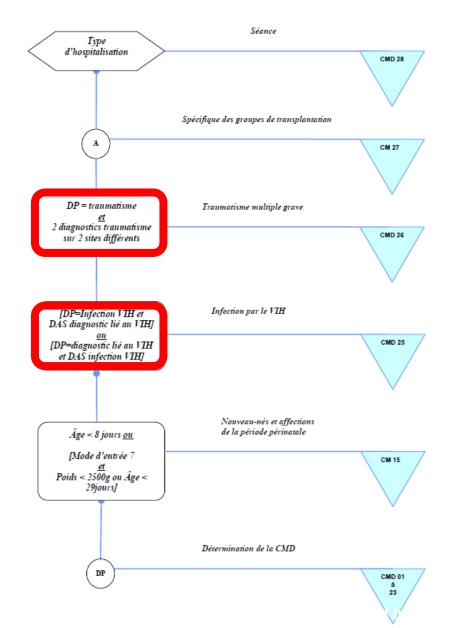
## Les diagnostics associés significatifs, porteurs de valeurs

- À travers (détaillé par la suite) :
  - Orientation vers une CMD
  - Orientation vers une racine de GHM
  - Orientation vers un GHM en 2/3/4 (CMA)
  - Orientation vers un GHM en B/C/D
  - Obtention du supplément de surveillance continue



#### ORIENTATION

vers les CATÉGORIES MAJEURES 27 et 15 vers les CATÉGORIES MAJEURES DE DIAGNOSTIC 1 à 14, 16 à 23, 25, 26 et 28



## Les DAS peuvent parfois orienter dans une CMD

(Volume 1 du Manuel des GHM)

La plupart des CMD : entrée par le

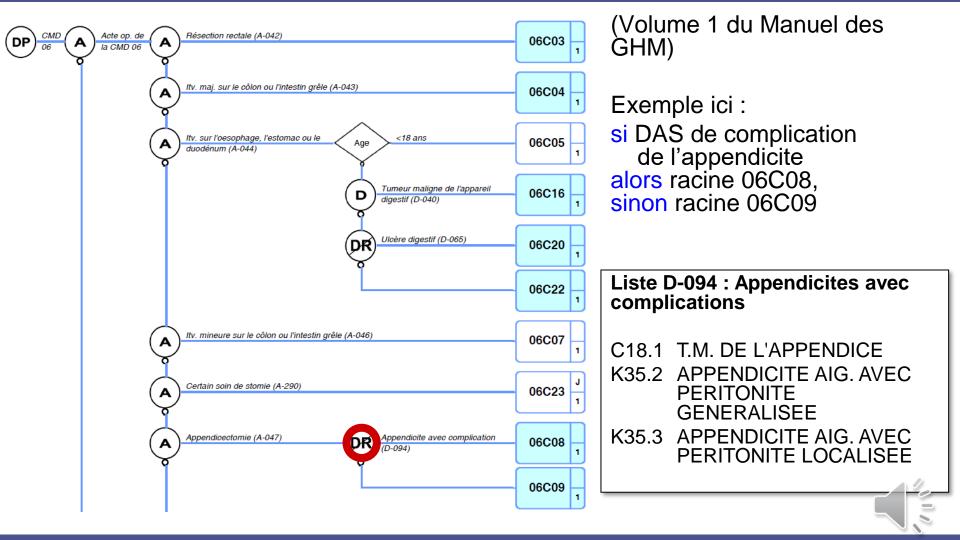
diagnostic principal

CMD 26: DAS de traumatisme notamment

CMD 25 : DAS lié au VIH notamment

<sup>\*</sup> parfois il s'agit dans le texte de « Catégories Majeures », CM

## Les DAS peuvent parfois orienter dans une racine de GHM



## DAS valorisés en Complications et Morbidités Associées (CMA)

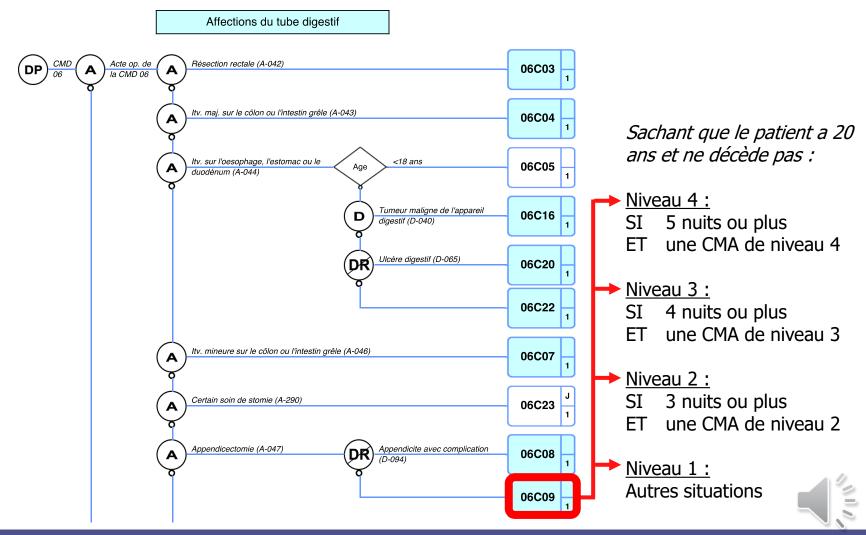
### Définitions :

- Ce sont certains DAS entraînant une augmentation significative du coût ou de la durée d'un séjour. IL FAUT DONC LES CODER AU MIEUX.
- Complications = en rapport avec la pathologie ou le traitement
  - ex : désunion de plaie opératoire
- Morbidité associée = pathologie intercurrente
  - ex : anémie falciforme connue chez un patient venant pour appendicite
- Abus de langage : « comorbidités associées »
  - pléonasme et omission du concept de complication
- Positions dans l'algorithme
  - Les CMA forment une ramification terminale, après affectation dans une racine de GHM
  - Les DAS correspondants ne sont pas les seules conditions d'accès à un GHM avec CMA (aussi : âge, décès, durée, âge gestationnel)
  - Certaines racines de GHM ne sont toutefois pas segmentés

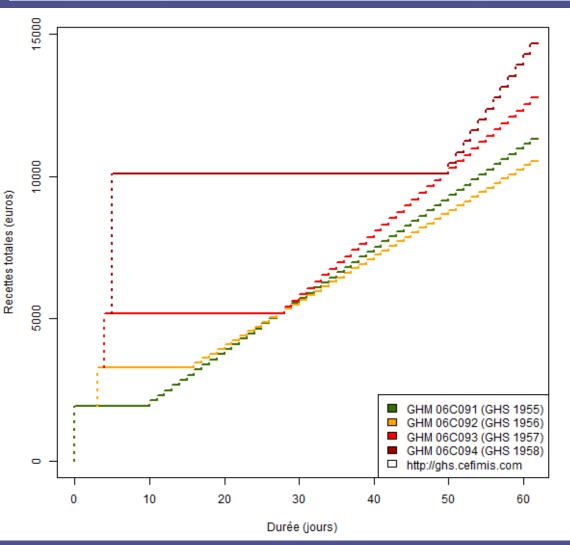
## Concrètement, effet conjoint de la durée et des CMA

- Exemple concret :
  - Patient venant pour une appendicite aiguë non compliquée (K35.9, diagnostic d'entrée en CMD 6)
  - On réalise une appendicectomie par voie iliaque (HHFA001, liste A047)
  - L'âge est de 20 ans
  - Le patient ne décède pas
- Testons toutes les durées (tarifs 2015 ex-DGF) :
  - Pas de séquelle notable
  - Avec une CMA de niveau 2 : G82.1 Paraplégie spastique
  - Avec une CMA de niveau 3 : G82.0 Paraplégie flasque
  - Avec une CMA de niveau 4 : G82.3 Tétraplégie flasque

### Concrètement, effet conjoint de la durée et des CMA



GHS	GHM	Libellé	Borne haute	Tarif	Tarif EXH
1955	06C091	App. non compliquées, niveau 1	9	1 930,13	180,78
1956	06C092	App. non compliquées, niveau 2	15	3 291,58	157,71
1957	06C093	App. non compliquées, niveau 3	27	5 184,11	222,81
1958	06C094	App. non compliquées, niveau 4	49	10 086,00	382,54



Recettes totales (€) en fonction de la durée de séjour (jours)

Tous les GHM 06C09\*, Appendicectomies non compliquées tarifs ex-DGF mars 2015 GHM v11g

→ Toutes les courbes sur http://ghs.cefimis.com

## Et dans le secteur privé lucratif (ex-OQN)?

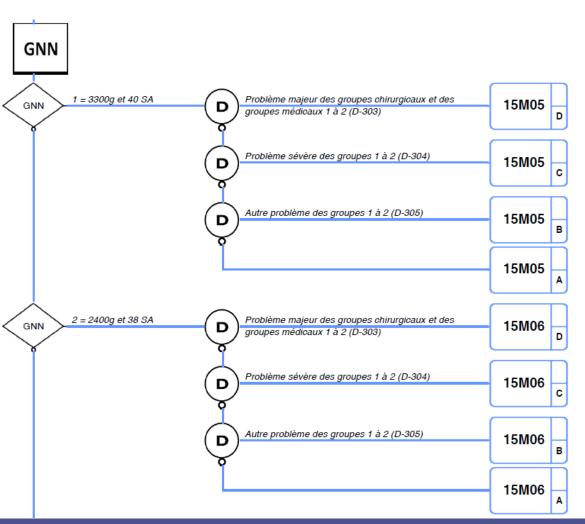
 La mécanique est similaire, mais les tarifs nettement plus faibles

GHS	GHM	Libellé	Tarif ex-	Tarif ex-
			DGF	OQN
1955	06C091	App. non compliquées, niveau 1	1 930,13	981,68
1956	06C092	App. non compliquées, niveau 2	3 291,58	1 496,15
1957	06C093	App. non compliquées, niveau 3	5 184,11	2 724,81
1958	06C094	App. non compliquées, niveau 4	10 086,00	4 294,44

- Les honoraires des praticiens sont payé en plus (et le praticien reverse des loyers et commissions à la clinique) :
  - L'acte HHFA001 du chirurgien : 176,79 €
  - L'acte HHFA001 de l'anesthésiste : 78,49 €
  - La radio de l'abdomen ZCQK002 : 19,95 €
  - L'examen anatomo-pathologique ZZQP188 : 33,60 €

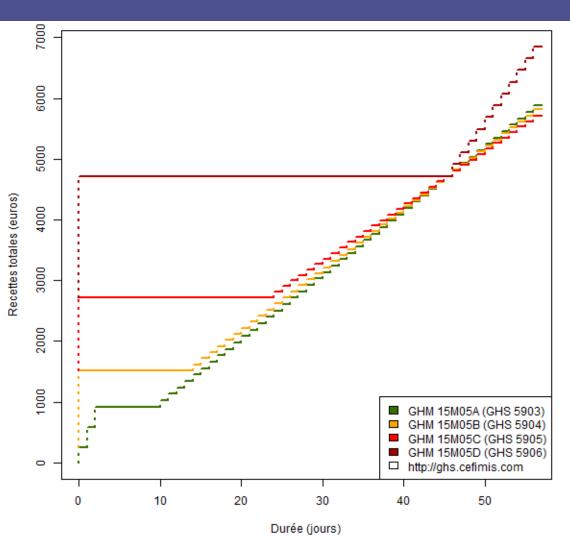


### Les GHM en B/C/D, des quasi-CMA



- Ramifications terminales ABCD : niveaux de complexité prévus ad hoc
- Pas aussi générique que les CMA

### Exemple d'une naissance d'enfant à l'hôpital



Recettes totales (€) en fonction de la durée de séjour (jours)

Tous les GHM 15M05\*, Nouveau-nés de 3300g et âge gestationnel de 40 SA et assimilés tarifs ex-DGF mars 2015 GHM v11g

→ Toutes les courbes sur http://ghs.cefimis.com

2020-04-17

### Les DAS pourvoyeurs de suppléments : supplément de surveillance continue

```
RUM effectué dans une UM avec autorisation Surveillance Continue
         * Patient sortant d'une UM autorisée Réanimation ou
Réa pédiatrique ou Réa néonatale avec sup +
         * ET = * âge < 18

* Un diag. marqueur L3 A8 2012
       * âge >= 18

* IGS2 sans âge >= 7

* Un diag. marqueur L1 A8 2012
         * IGS2 sans âge >=15
         * Un acte marqueur L2 A8 2012 dans ce RUM ou le précédent
```

 Les DAS peuvent participer à l'obtention d'un supplément de surveillance continue

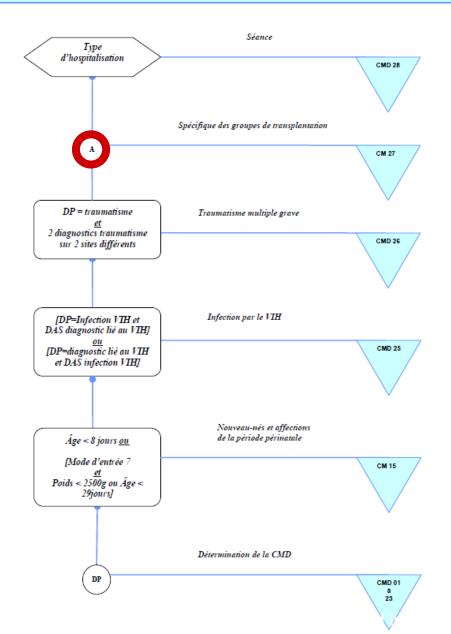
## Les actes CCAM, porteurs de valeurs

- À travers (détaillé par la suite) :
  - Orientation vers une CMD
  - Orientation vers une racine de GHM
  - Obtention du supplément de réanimation ou de surveillance continue



#### ORIENTATION

vers les CATÉGORIES MAJEURES 27 et 15 vers les CATÉGORIES MAJEURES DE DIAGNOSTIC 1 à 14, 16 à 23, 25, 26 et 28



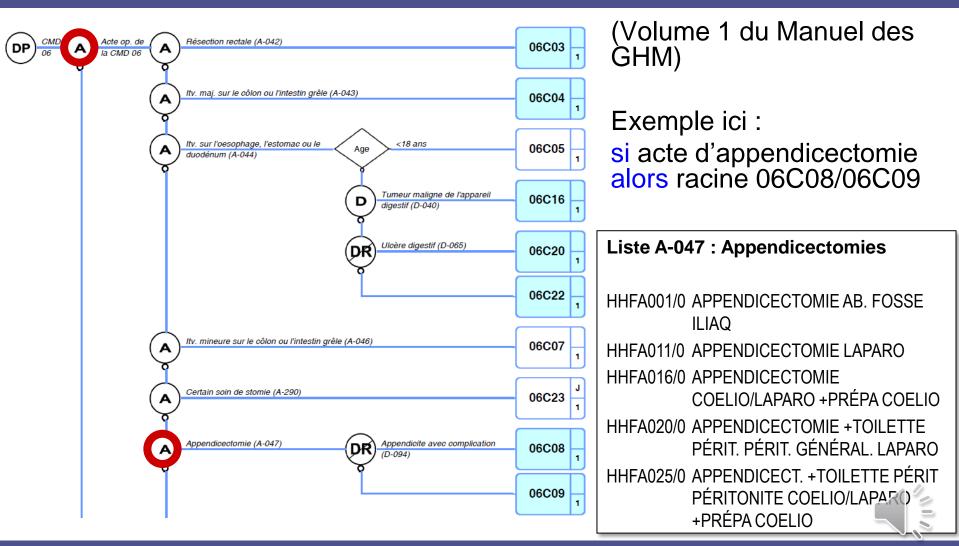
## Les actes peuvent parfois orienter dans une CMD

(Volume 1 du Manuel des GHM)

La plupart des CMD : entrée par le diagnostic principal CMD 27 : Acte CCAM de transplantation d'organe

<sup>\*</sup> parfois il s'agit dans le texte de « Catégories Majeures », CM

## Les actes « classants » peuvent orienter dans une racine de GHM



## Les actes pourvoyeurs de suppléments

### Réanimation

# RUM effectué dans une UM avec autorisation Réanimation Un acte marqueur L1 A7 (2011) ex : ABQP001 Surveillance continue de la pression intracrânienne, par 24h. 3 itér. acte marqueur L2 A7 (2011) ex : GLLD 015 Ventilation mécanique intra trachéale avec PEP <6 et FiO2 ≤ 60 %, par 24h. IGS2 ≥ 15 Âge < 18

### Surveillance continue

```
RUM effectué dans une UM avec autorisation Surveillance Continue

* Patient sortant d'une UM autorisée Réanimation ou Réa pédiatrique ou Réa néonatale avec sup +

* âge < 18

* Un diag. marqueur L3 A8 2012

* âge >= 18

* ET  

* IGS2 sans âge >= 7

* Un diag. marqueur L1 A8 2012

* IGS2 sans âge >= 15

* Un acte marqueur L2 A8 2012 dans ce RUM ou le précédent
```

 Les actes peuvent participer à l'obtention d'un supplément de réanimation ou surveillance continue

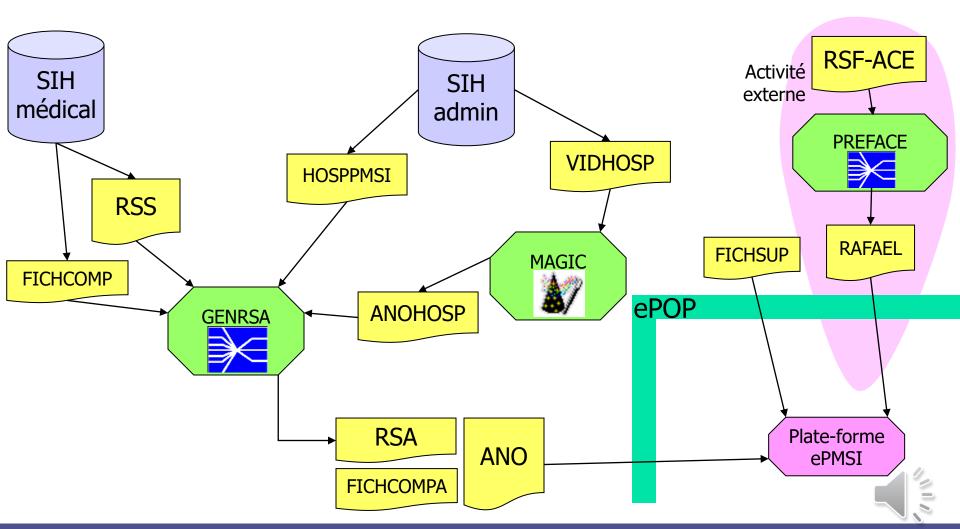
### Morale de l'histoire

- Le diagnostic principal :
  - Rôle crucial
  - Rarement modifié par le DIM
  - Impossible d'avoir une attitude générique simple, hormis quelques règles expertes de permutation
- Certains DAS et certains actes :
  - Potentiel d'augmentation de valeur
  - Mais pas systématique : très contextuel
- Conséquences :
  - Nous parlerons désormais de « DAS ou actes potentiellement valorisants » et pas forcément de CMA
  - But du contrôle qualité automatisé : pouvoir <u>rajouter</u> des codes, sans se poser d'autre question, avec ou sans impact sur la valeur
  - Conséquence : nécessité d'une simulation de groupage

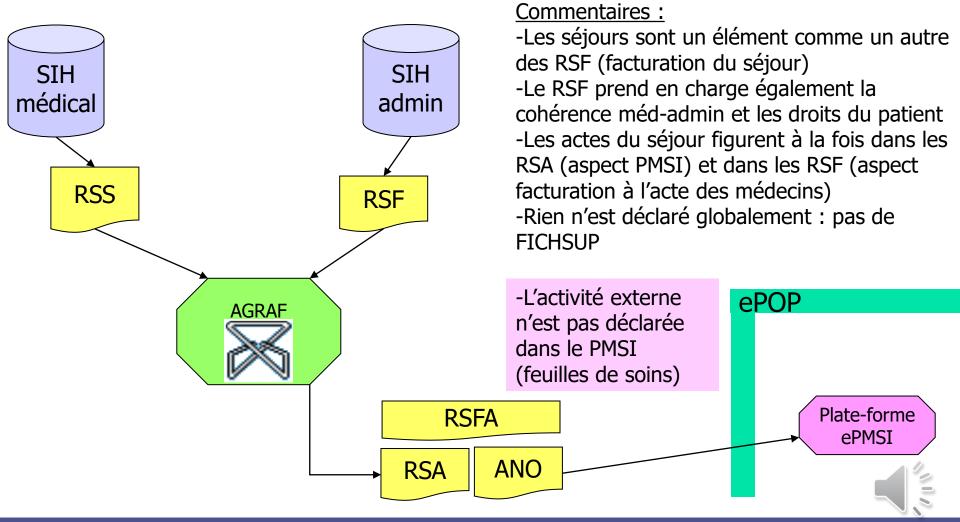
# Description of the French PMSI files: de facto, an interoperability standard!



# Émission des fichiers MCO (modèle transitoire ex-DGF)



# Émission des fichiers MCO (secteur ex-OQN, modèle cible)

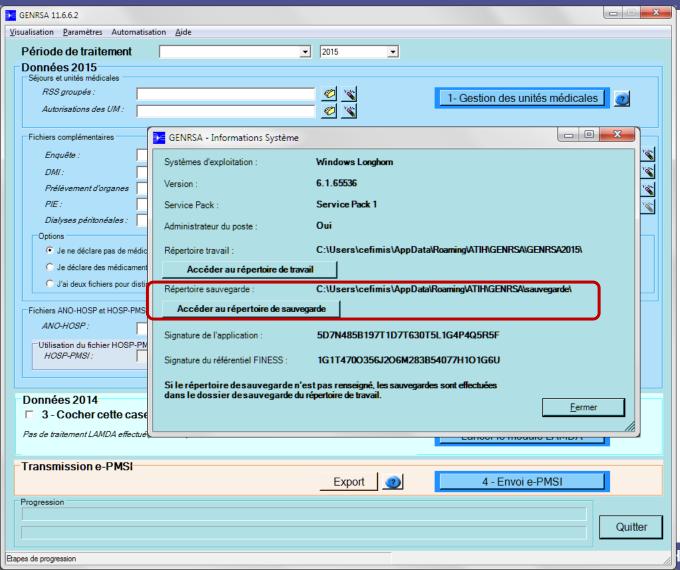


#### Présentation de GENRSA

- En MCO ex-DGF, programme permettant à la fois le chaînage, le groupage et l'anonymisation des RUM
- Téléchargement :



# GENRSA: répertoires de sauvegarde



Menu Aide →
Infos système et
paramétrage

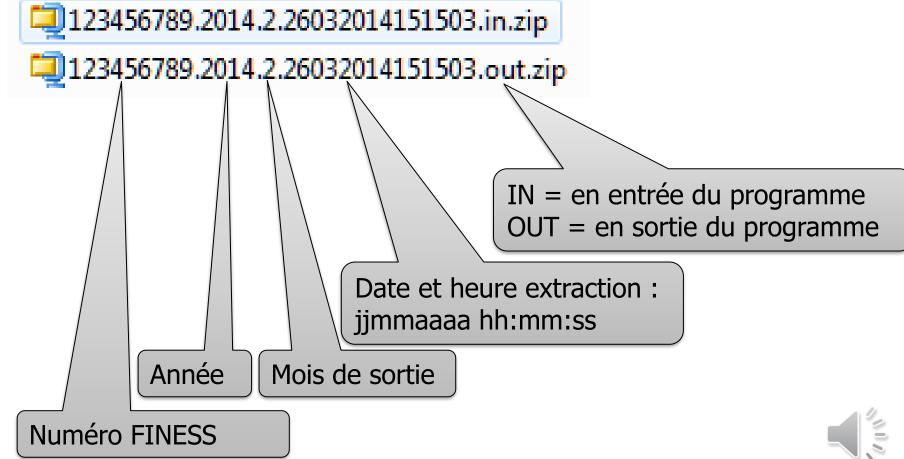
Menu Aide → Manuel d'utilisation



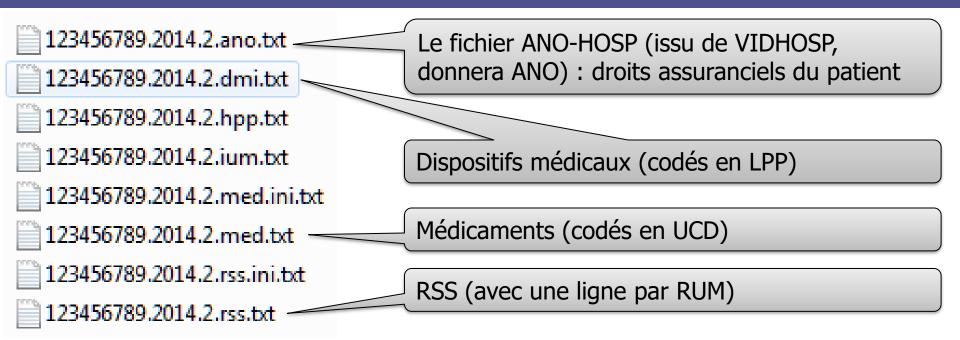
lealthcare

### Fichiers de sauvegarde de **GENRSA**

Fichiers IN et OUT, zippés :



#### Fichier IN de GENRSA



#### Nous n'utiliserons pas ces fichiers car :

- On retrouve les mêmes informations en sortie du programme
- Le format RSA, avec une ligne par séjour, est plus pratique
- On retrouve aisément les identifiants!

123456789,2014,2,ano 123456789.2014.2.chainage.err.txt 123456789.2014.2.chainage.log.txt 123456789.2014.2.comp.ansi.log.txt 123456789,2014,2,comp,log,txt 123456789.2014.2.diap 123456789,2014,2,dif.txt 123456789.2014.2.dmip .23456789.2014.2.doublonsnorss.log.csv 123456789.2014.2.eng 23456789.2014.2.ium 123456789.2014.2.leg 123456789.2014.2.log.txt 123456789.2014.2.med 123456789.2014.2.medatu 23456789.2014.2.medthrombo

123456789.2014.2.orp

123456789.2014.2.pie

123456789.2014.2.pip

123456789.2014.2.porg

123456789.2014.2.sejnonautor.txt

123456789.2014.2.rsa

123456789.2014.2.sta

alctl mt2a.detail.log.csv

ctl mt2a.synth.log.txt

123456789,2014,2,tra,txt

Fichier OUT de GENRSA

Listing complet des fichiers:

(diapo suivante : un extrait)



### Fichier OUT de GENRSA (extrait)

- 123456789.2014.2.ano
- 123456789.2014.2.dmip
- 123456789.2014.2.med
- 123456789,2014,2,medatu
- 123456789,2014,2,medthrombo
- 123456789.2014.2.rsa
- 123456789.2014.2.tra.txt

Le fichier ANO : droits assuranciels du patient et numéro unique du patient !! (Une ligne par séjour)

Dispositifs médicaux (codés en LPP)

Médicaments (codés en UCD) :

- Molécules onéreuses
- Médicaments sous ATU
- Médicaments anti-thrombotiques

Fichier de RSA (avec une ligne par séjour)

Fichier TRA (avec une ligne par séjour)



### Formats des fichiers : décrits dans la notice de GENRSA

	Manuel d'utilisation de GENRSA Version 11.6.X Janvier 2015		
3	LE MODULE LAMDA		
Les	Les fichiers d'entrée de traitement de LAMDA		
3.2	Lancement du module LAMDA		
3.3	L'interface principale du module LAMDA35		
Réc	upération des fichiers de signature		
3.4	Le traitement des séjours et fichiers complémentaires		
	vération des données M12 2014 mises à jour		
	vix des fichiers de sauvegarde de M12 2014 d'origine		
4 MODE	TRANSMISSION ET TRAITEMENT DES DONNÉES SUR E-PMSI : E D'EMPLOI		
4.1	GENRSA 43		
4.2	e-PMSI		
5	L'UTILISATION D'AGORA48		
6	ANNEXES52		
6.1	Configuration minimale requise		
6.2 RSA)	Sauvegarde du fichier de RSS à la source du fichier de RSAc (ou de , sécurisation des données et fonctionnement du contrôle externe 54		
6.3	Formats de RSS groupé 116		
6.4	Formats de RSS groupé 11760		
6.5	Format ANO-HOSP 2014 64		
	6.6 Format ANO-HOSP 2015		
6.7 GENI	Format d'importation du fichier d'information des UM (utilisé par RSA)74		
6.8	(*) Codage Type d'autorisation de l'unité médicale		
6.9	Format 2008 du fichier HOSP-PMSI78		
6.10	Formats des fichiers complémentaires		
	Page 3/131		
0820 77 1234 - suppo https://applis.atih.san	ort@atih.sante.fr Agence technique de l'information sur l'hospitalisation (ATIH)		

Fichiers en sortie

Fichiers en entrée

Version 11.6.X 6.11 Format de RSA groupé en CCAM (222) 80 6.12 Fichier de chaînage (ANO) 88 6.13 Fichiers LEG et STA 92 Format du Fichier LEG 92 Liste complète des codes erreurs. 92 97 Liste détaillée des codes erreurs de contrôles Liste détaillée des codes erreurs de groupage. 119 Liste détaillée des codes erreurs d'implémentation 122 Format du fichier STA 124 6.14 Formats des fichiers de sortie suivi par patient supplémentaires... 125

Manuel d'utilisation de GENRSA

6.15 Problèmes connus

0820 77 1234 - support@atih.sante.fr https://applis.atih.sante.fr/agora

117, bd Marius Vivier Merle 693 Téléphone : 04 37 91 33 10 - 1

### Que trouve-t-on d'autre dans GENRSA? Des formats de fichiers

- C:\Program Files (x86)\ATIH\GENRSA\sys\fmt
- Fichiers décrivant comment découper certains fichiers du PMSI
- Moins complet que les notices papier, mais plus de fichiers décrits
- Exemples : certains champs du RSS, certains champs du RSA, médicaments, DMI, ANO, TRA...
- Exemple du fichier TRA :

  Le numéro RSA sur 10 caractères

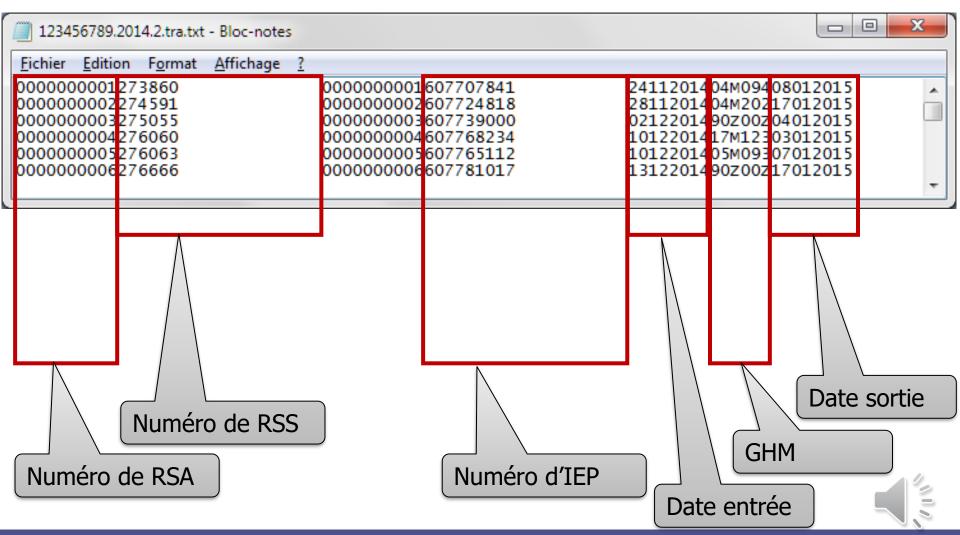
  Le numéro RSS sur 20 caractères

  Le numéro IEP sur 20 caractères

#### Un mot sur le fichier TRA

- Décrit (1 ligne par séjour, correspondance ligne à ligne avec le RSA et l'ANO) :
  - Numéro de RSA
  - Numéro de RSS
  - Numéro d'IEP / IAS / NAS
  - Autres : dates, GHM
- Présent dans le OUT de GENRSA, mais supprimé par le module ePOP avant l'envoi sur ePMSI
- Utilisé par LEDDA pour « casser » l'anonymat

### Exemple de fichier TRA (fictif)



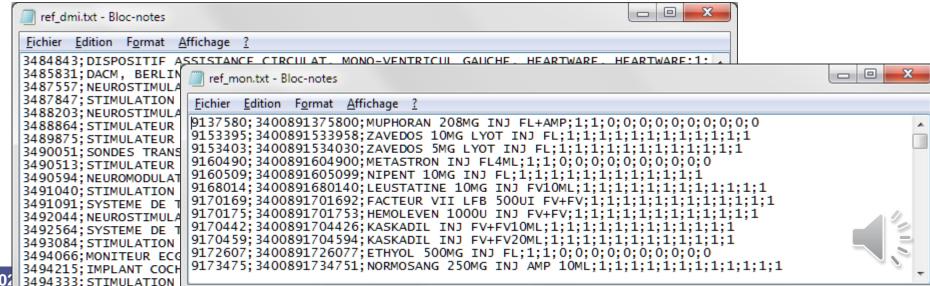
### Fichier ANO: contient notamment le numéro ANO du patient

- Il existe dans les 5 champs un numéro unique du patient, partout en France
- Calcul du numéro ANO du SNIIR niveau 1 :
  - Calculé en utilisant le numéro NIR (SS) de l'assuré (parent), date de naissance et sexe du bénéficiaire (pas le rang de naissance)
  - Supposé unique pour un patient, mais :
    - Change lorsque le patient change de régime (majorité, conjoint à charge, MSA, militaires...)
    - Deux jumeaux de même sexe ont le même numéro ANO jusqu'à leur majorité
  - Présent dans les fichiers ANO du OUT.ZIP (comparables entre établissements, ce qui est interdit)
- Numéro ANO du SNIIR niveau 2 :
  - Crypté après transmission
  - Différent, mais identité conservée entre établissements, entre champs
  - Disponible dans la Base Nationale du PMSI
- Attention : malheureusement, l'ANO est calculé sur la base du NIR de l'assuré, pas du bénéficiaire...



### Que trouve-t-on d'autre dans GENRSA? Des tables de référence

- C:\Program Files (x86)\ATIH\GENRSA\sys\ref
- Fichier de terminologies :
  - LPP Liste des Produits et Prestations de la SS, pour les dispositifs médicaux implantables
  - UCD Unités Communes de Dispensation, pour les médicaments (molécules onéreuses, médicaments sous ATU, médicaments thrombolytiques)



# D'autres programmes à explorer...

- MCO :
  - GENRSA pour l'ex-DGF
  - AGRAF pour l'ex-OQN
  - LEDDA pour le contrôle qualité DATIM
- SSR:
  - GENRHA pour l'ex-DGF
  - AGRAF-SSR pour l'ex-OQN

- PSY: Pivoine
- Soins externes :
  - PREFACE pour l'ex-DGF
  - Rien en ex-OQN
- HAD:
  - PAPRICA-DGF pour l'ex-DGF
  - PAPRICA-OQN pour l'ex-OQN
  - LEDDA pour le contrôle qualité DATIM

NB: et MAGIC pour toutes les applications ex-DGF

# Présentation des données en hospitalisation : exemple de fichiers décrivant un séjour hospitalier MCO ex-DGF versus ex-OQN

#### Données (simplifiées):

- 1234 : numéro de séjour anonyme
- K359 : un code diagnostique
- HHFA001 : un code d'acte
- 1955 : un numéro de GHS (correspondant à un tarif)
- DMIXXX : un code de LPP
- MONXXX : un code UCD
- SS: un régime d'assurance maladie
- 100%: taux de prise en charge du patient

#### Représentation ex-OQN

(hormis honoraires consultations & suivi)

Fichier de RSA (codage)

1234...1955...K359...HHFA001...¶

Fichier de RSFA (facture) A...1234...\$S...100...¶
B...1955...¶
P...DMIXXX...¶
H...MONXXX...¶
M...HHFA001...¶

#### Représentation ex-DGF

Fichier de RSA (valorisation) 1234...1955...K359...HHFA001...¶

Fichier ANO (valorisation)

1234...\$S...100...¶

Fichcompa LPP (valorisation)

1234...DMIXX...100...¶

Fichcompa UCD (valorisation)

1234...MONXX...100...¶

### Quality control procedures

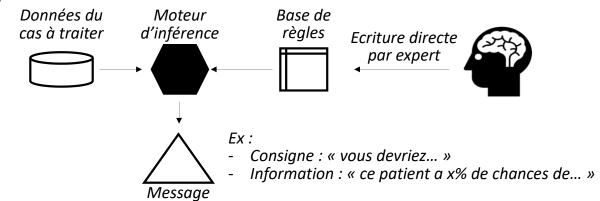


# Organisations du contrôle / optimisation

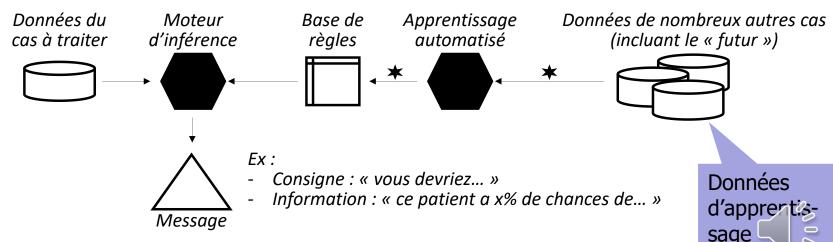
- Contrôle aléatoire
  - Anciennement en ex-DGF
- Contrôle chronologique
  - Actuellement en ex-OQN, car facturation au fil de l'eau
- Contrôle thématique
  - Actuellement en ex-DGF
  - Ciblé sur des problèmes connus
  - Ciblé sur des anomalies détectées par des règles

# Two levels of artificial intelligence

Niveau 1 : exécuter



Niveau 2 : apprendre et réappliquer



# Trois familles de règles de contrôle qualité automatisé

- Règles expertes : IA niveau 1
  - Ecrites à la main
  - Permettent des opérations complexes : permutations, ajouts, suppression
  - Exemple : si « suture immédiate au décours de l'accouchement » et pas « accouchement », alors remplacer l'acte par « plastie (…) »
- Règles maladies chroniques (irréversibles) : IA niveau 1
  - Une règle unique, alimentée par une liste de maladies chroniques
  - Simple rajout de DAS
  - Exemple : si « Alzheimer » les mois précédents et pas cette fois-ci, alors reconduire le code
- Règles issues de machine learning : IA niveau 2
  - Découverte automatique d'associations
  - Rajout de codes lorsqu'ils manquent alors que c'est rare
  - Exemple : si « appendicite » et pas « transfert », alors « appendicectomie »



# Exemple de règle experte (IA niveau 1)

Informations

Homme de 74 ans Séjour de 6 jours

DP: R509 Fièvre, sans précision

DAS: B349 Infection virale, sans précision

Anomalie:

Code symptôme en DP (R509), étiologie en DAS (B349)

Proposition: Inverser le DP et le DAS

Avant : 18M042

Fièvres d'étiologie indéterminée, âge supérieur à 17 ans, niveau 2. Après : 18M112

Autres maladies infectieuses ou parasitaires, niveau 2

2849 €

**3351 €** soit **+502€** 

2020-04-17

**Proposition** 

### Exemple de règle Maladie chronique (IA niveau 1)

Homme de 43 ans Séjour de 9 jours

DP: N10 Néphrite tubulo-interstitielle aiguë

Anomalie:

Précédent séjour : antécédent de Sida (B24+1) il y a 0.8 mois

Proposition: Ajouter **B24+1** en DAS

Avant : 11M041

Infections des reins et des voies urinaires, âge supérieur à 17 ans, niveau 1

Après : 25M02B

Maladies dues au VIH, avec une seule complication infectieuse.

1223€

**5135** € soit **+3912**€

# Exemple de règle Machine learning (IA niveau 2)

Femme de 79 ans Séjour de 5 jours

DP: A46 Erysipèle

DAS: D689 Anomalie de la coagulation, sans précision

MON: 9225918 (UCD) Factane (facteur VIII)

L'explication sera généralement absente

Anomalie: Proposition:

Administration de **Factane** sans indication thérapeutique Ajouter **D66** en DAS (carence héréditaire en facteur VIII)

Avant : 09M052

Lésions, infections et inflammations de la peau et des tissus souscutanés, âge supérieur à 17 ans, niveau 2

Après : 09M053

Lésions, infections et inflammations de la peau et des tissus souscutanés, âge supérieur à 17 ans, niveau 3

3064€

**4319 €** soit **+1254 €** 

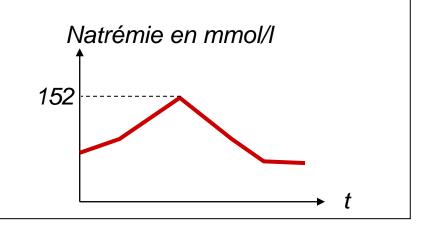
**Proposition** 

# Exemple de règle utilisant les résultats de biologie médicale (experte ou *machine learning*)

nformations

Femme de 62 ans Séjour de 8 jours

DP: M179 Gonarthrose



**Proposition** 

Anomalie : Hypernatrémie > 150mmol/l visible dans les résultats de biologie Proposition : Ajouter E8700 en DAS (hypernatrémie > 150mmol/l)

Avant : 08M101

Maladies osseuses et arthropathies spécifiques, niveau 1

Après : 08M102

Maladies osseuses et arthropathies spécifiques, niveau 2

1516€

**5303** € soit **+1987**€

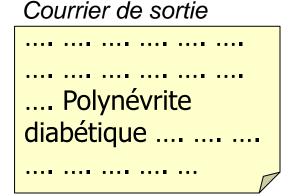


#### Exemple de règle utilisant les courriers médicaux (experte ou *machine learning*)

nformations

Femme de 52 ans Séjour de 3 jours

DP: E109 DID sans complication



Proposition

Anomalie: Polynévrite diabétique dans le courrier, DID non compliqué codé. Proposition: Coder E104 en DP et G632 en DAS (polynévrite diabétique)

Avant: 10M021

Diabète, âge supérieur à 35 ans, niveau 1

Après: 01M111

Affections des nerfs crâniens et rachidiens, niveau 1

1863 €

2052 € soit +189 €



### Les règles expertes (IA niveau 1)

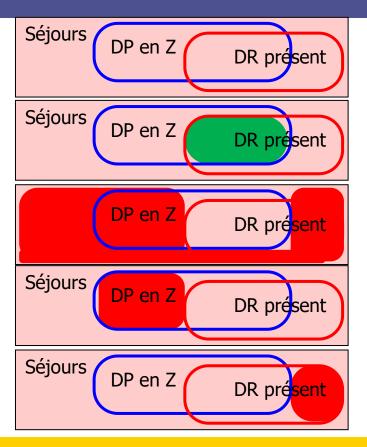
Exemples: Nestor, Datim, (2000), Ovalide





### Règles expertes

- Exemple : « lorsque le DP commence par la lettre Z, on peut coder un DR »
- Situation normale : séjours avec DP en Z et DR présent
  - Règle fausse : alerte si pas (DP en Z et DR)
  - Règle fausse : alerte si DP en Z et pas de DR
- Situation anormale : séjour avec DR alors que le DP n'est pas en Z
  - Règle exacte : alerte si DR alors que le DP n'est pas en Z



Conclusion : le texte désigne généralement des situations normales (possibles / probables). Avant d'implémenter une règle, il faut en déduire des situations anormales (interdites). Couvrent très partiellement le domaine, mais sans erreur.

#### Au bilan

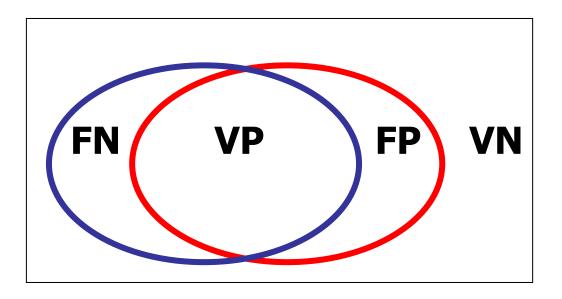
- Consigne = situation normale
  - « lorsque le DP commence par la lettre Z, on peut coder un DR »
- Situation anormale :
  - DR présent alors que le DP ne commence pas par la lettre Z
  - Solution proposée : déplacer le DR en DAS
- Règle :
  - si DR≠"" et GAUCHE(DP, 1)≠"Z"
     alors alerte("Déplacer le DR en position de DAS")

# Démarche itérative d'amélioration des règles

- Notre cerveau n'est pas surpuissant : impossible de prévoir tous les cas de figure
- Démarche itérative :
  - Implémenter une règle
  - Revue « experte » d'un certain nombre de cas déviants
  - Description des motifs d'échecs de la règle : faux positifs : bruit = 1-VPP = FP/(VP+FP)
  - Ajout de conditions d'exclusion, pour réduire le bruit

Ensemble bleu : Séjour normal/anormal Ensemble rouge : Règle positive (alerte)/négative

N/A R+/R-



FN = R-  $\cap$  A VP = R+  $\cap$  A FP = R+  $\cap$  N VN = R-  $\cap$  N



N = Séjour normal

A = Séjour anormal

R+ = règle positive (alerte)

R- = règle négative (silence)

	A	N
R+	# VP	# FP
R-	# FN	# VN

 On regarde toujours les proportions de bien classés dans une colonne ou une ligne :

	А	N
R+	VP	FP
R-	FN	VN

Valeur prédictive positive

$$= VPP$$

$$= P(A \mid R+)$$

$$= VP / (VP+FP)$$

Valeur prédictive négative

$$= P(N \mid R-)$$

$$= VN / (VN+FN)$$



 On regarde toujours les proportions de bien classés dans une colonne ou une ligne :

	А	N
R+	VP	FP
R-	FN	VN

```
Valeur prédictive positive
= VPP
= P(A | R+)
= VP / (VP+FP)
```

Valeur prédictive négative = VPN = P(N | R-) = VN / (VN+FN)

Sensibilité = Se  
= 
$$P(R+ | A)$$
 Spécificité = Sp  
=  $VP / (VP+FN)$  =  $P(R- | N)$   
=  $VN / (VN+FP)$ 



#### Validation d'une règle :

En pratique on cherchera uniquement à augmenter la VPP (diminuer le bruit) :

	Α	N	Valeur prédictive positive
R+	VP	FP	= VPP = P(A   R+) = VP / (VP+FP)
R-	FN	(VN)	= VP / (VP+FP)

- Ajouter des exclusions => déplacer les effectifs selon la flèche bleue
- Pourquoi diminuer le bruit / augmenter la VPP ?
  - on ne peut pas tout contrôler!
  - utiliser au mieux le temps dédié au contrôle



#### Exemple

- Consigne :
  - « les patients avec appendicite doivent avoir une appendicectomie »
- Situation anormale:
  - appendicite sans appendicectomie
- Règle :
  - si appendicite et pas d'appendicectomie, alors alerte
- Revue de cas :
  - Nombreux faux positifs
  - Notamment patients transférés ou décédés
- Nouvelle règle :
  - si appendicite et <u>pas d'appendicectomie et pas de</u> transfert et <u>pas de</u> décès, <u>alors alerte</u>



# Les règles de maladies chroniques Un type particulier de règle experte (IA niveau 1)



### Règles de maladies chroniques

### Principe :

- Certaines maladies sont irréversibles ou lentement réversibles
- Et entraînent très souvent des soins, quel que soit le motif du séjour
- Bien entendu, il faudra toujours vérifier

### Exemples :

- Par définition, les séquelles et amputations (absence acquise d'un membre)
- Les présences de prothèses (présence de valve cardiaque métallique)
- Les défaillances irréversibles d'organes (insuffisance rénale chronique)
- Certaines démences (Alzheimer)
- Certaines infections (VIH, hépatite C)
- Des états d'évolution lente (obésité morbide)



### Construction des règles

- Attention à la constitution de la liste de codes!
- Maladies chroniques groupées par catégorie :
- Principe :
  - si un patient valide une catégorie puis la revalide (même avec un code différent), OK
  - Si un patient valide puis ne revalide pas, on reconduit le dernier code observé.
- Implémentation : lier par patient (ANO)

alcoolisme antécédent personnel ou familial auto-immunité maladie cardio-vasculaire chronique anomalie chromosomique maladie dermatologique chronique diabète diabète compliqué maladie digestive chronique présence d'un dispositif maladie endocrino chronique dépendance ou handicap maladie hémato chronique maladie infectieuse chronique tumeur maligne tumeur maligne abdominale tumeur maligne endocrino cancer gynécologique cancer hématologique

tumeur maligne neuro tumeur maligne ORL tumeur maligne osseuse tumeur maligne thoracique cancer urologique maladie locomotrice chronique maladie métabolique chronique maladie neurologique chronique obésité état pré-cancéreux hématologique maladie psychiatrique chronique maladie rénale chronique maladie respiratoire chronique séquelle tabagisme trouble du rythme toxicomanie retard du développement tumeur suspecte valvulopathie

# Règles créées par *machine learning* (IA de niveau 2)



# Machine Learning IA de niveau 2

- Les règles ne sont pas écrites par un humain, mais déduites automatiquement des associations fréquentes par une machine
- Avantages :
  - Valides de fait dans l'échantillon
  - Générées automatiquement, envisagent des milliers de situations différentes (y compris imprévisibles!)
  - Incluent de facto plusieurs critères de segmentation ou d'exclusion
  - Données hétérogènes mélangées : CIM10, CCAM, UCD, LPP, etc.
- Inconvénients :
  - Que pour les situations suffisamment fréquentes
  - Pas de validation experte : toujours vérifier les cas
  - Risque d'association fortuite (pouvoir prédictif conservé mais pas de validité théorique)
  - Aucune « explication » de l'association



### Confiance et VPP d'une règle

- Exemple de règle apprise :
  - Si Appendicite et pas de transfert alors 85% d'appendicectomie
  - 85% est la <u>confiance</u> : probabilité de l'appendicectomie lorsque les conditions sont remplies
- Règle de contrôle issue de la précédente :
  - Si Appendicite et <u>pas de</u> transfert et <u>pas d</u>'appendicectomie alors ajouter appendicectomie
  - La <u>VPP</u> est le succès de cette nouvelle règle, en revoyant les cas à dire d'expert!
- De notre expérience (heuristique) :
  - Si confiance > 80%, alors VPP > 40%
  - NB : en théorie, si codage parfait, VPP=0 quelle que soit la confiance
  - Conséquence : ignorer d'emblée les règles avec confiance < 70%</li>

# Procédé de construction des règles (1)

- Extraction de caractéristiques (features extraction) :
  - Feature extractor d'information administrative
    - Ex : « entrée par les urgences » vaut 1 si le mode d'entrée vaut 8 et la provenance vaut 5, ou si le motif de non facturation vaut 4.
  - Feature extractor de diagnostics
    - Exemple : 'diag1117' (appendicites) vaut 1 si on rencontre au moins un des codes CIM10 suivants : K35\*, K36, K37, K38\*
  - Feature extractor d'actes diagnostiques ou thérapeutiques
    - Exemple: 'acte0684' (appendicectomies) vaut 1 si on rencontre au moins un des codes CCAM suivants: HHFA001, HHFA011, HHFA016, HHFA020, HHFA025
  - Feature extractor de médicaments (codes UCD : molécules onéreuses, médicaments sous ATU, médicaments thrombolytiques)
  - Feature extractor de dispositifs médicaux implantables (codes LPP)

# Procédé de construction des règles (1)

Résultat de l'extraction de cadactéristiques :

Type de codes	Nombre de catégories mappées	Nombre de codes utilisés
Actes (CCAM)	1 340	8 451
Diagnostics (CIM10)	2 477	39 783
DMI (LPP)	60	1 851
Médicaments (UCD)	181	842
Information administrative	17	(sans objet)
Total	4 075	50 927

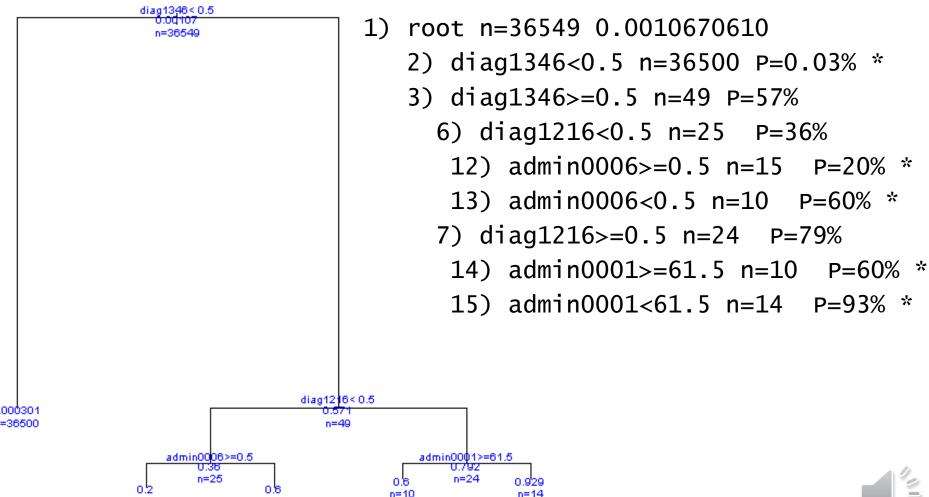


# Procédé de construction des règles (2)

- Puis machine learning supervisé :
  - Tour à tour, on tente de prédire chaque variable binaire de diagnostic ou d'acte par toutes les autres variables
- Puis filtrage des règles :
  - On supprime toutes les règles avec confiance < 80%</li>
- Puis transformation
  - des règles : A & B → C
  - En règles de contrôle : A & B & non-C → alerte, ajouter C
- Puis détection des cas :
  - On détecte les cas non-conformes (A & B & non-C)
- Puis filtrage des cas :
  - Simulation de groupage, calcul du gain potentiel
  - Elimination des séjours sans gain
  - Tri par {gain potentiel \* confiance de la règle} décroissant
- Puis revue un à un des dossiers



### Exemple de sortie



n = 15

n = 10

### Exemple de règle XML

```
<rule>
      <condition>
            <field>dmi0004</field>
            <operator><![CDATA[>=]]></operator>
            <ref>0.5</ref>
      </condition>
      <condition>
            <field>acte1215</field>
            <operator><![CDATA[< ]]></operator>
            <ref>0.5</ref>
      </condition>
      <condition>
            <field>diag2160</field>
            <operator><![CDATA[>=]]></operator>
            <ref>0.5</ref>
      </condition>
      <effect>
            <field>acte1218</field>
            <confidence>0.9815</confidence>
      </effect>
</rule>
```

#### Si on a un DMI

```
3301328 / 3304924 / 3305817 / 3306231 / 3312355 / 3314839 / 3317890 / 3322454 / 3326498 / 3341865 / 3343232 / 3347900 / 3350930 / 3355809 / 3365280 / 3373500 / 3396062 / 3399480
```

#### Et si on n'a pas d'acte

```
NEDA002 / NEKA001 / NEKA002 / NEKA003 / NEKA004 / NEKA005 / NEKA006 / NEKA007 / NEKA008 / NEKA009 / NEKA013 / NEKA015 / NEKA019 / NEKA022 / NELA001 / NELA002 / NFKA003 / NFKA004 / NFKA005 / NFLA002 / NFLA003 / NFLA002 / NFLA006
```

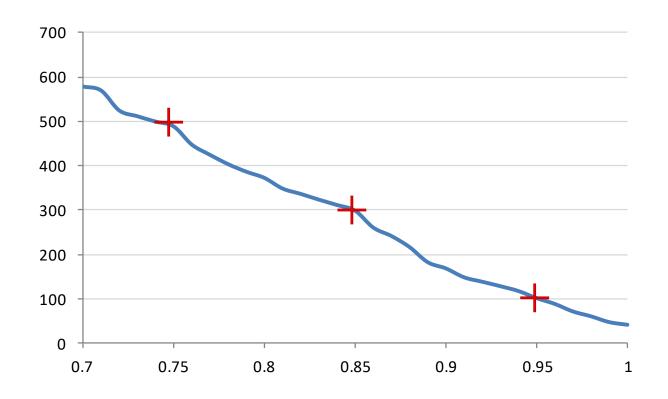
#### Alors on a 98% de chances d'avoir un acte

NEKA010 / NEKA011 / NEKA012 / NEKA014 / NEKA016 / NEKA017 / NEKA018 / NEKA020 / NEKA021 / NEMA018

### Et si on a un diagnostic

S720 / S7200 / S721 / S7210 / S722 / S7220

# Nombre de règles selon le seuil de confiance



Nombre de règles dont la confiance est supérieure ou égale à un seuil donné (exemple base nationale PMSI)



# Métarègles pour limiter le bruit sur la prédiction d'actes classants

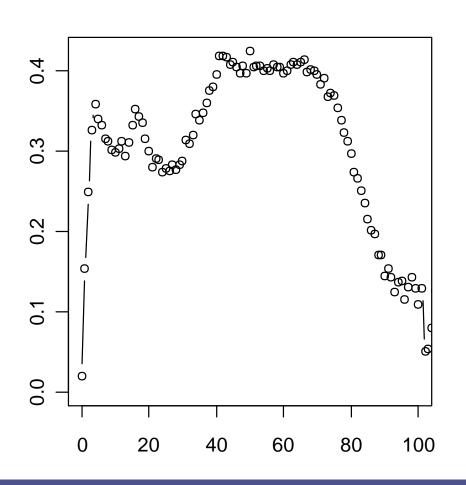
Métarègle = règle au sujet des règles

Autrement dit, super-règles permettant de modifier le comportement d'une règle. Ici pour diminuer le bruit, augmenter la VPP.

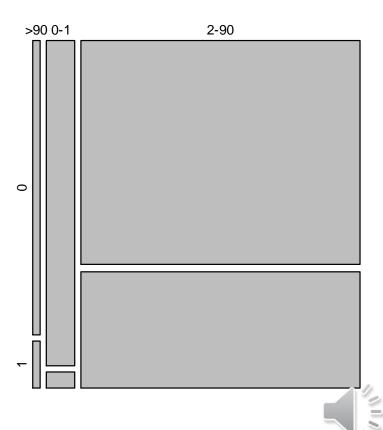


# La probabilité d'avoir un acte classant en fonction de l'âge

#### acte classant

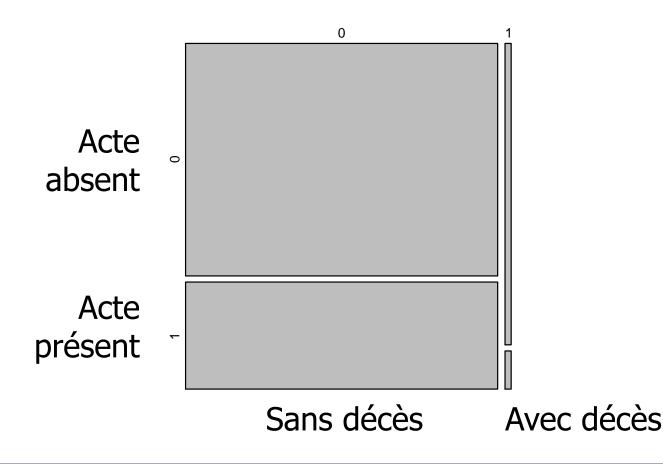


#### acte classant



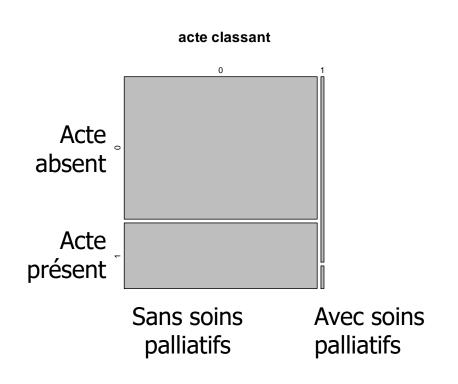
# La probabilité d'avoir un acte classant en fonction du décès

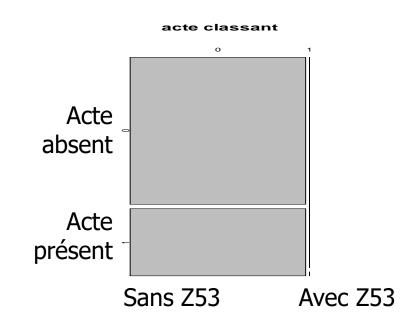
#### acte classant





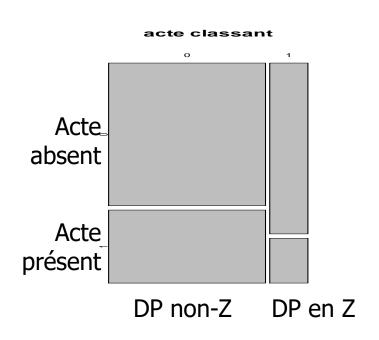
### La probabilité d'avoir un acte classant en fonction de certains DAS

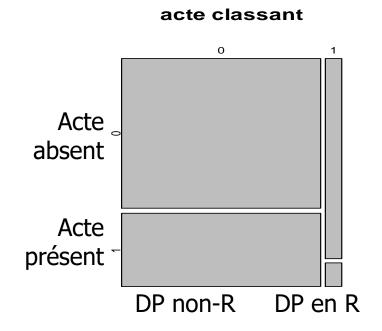






# La probabilité d'avoir un acte classant en fonction du DP

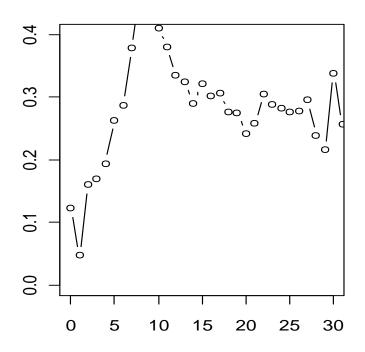






### La probabilité d'avoir un acte classant en fonction de la DS, en cas de transfert uniquement

#### acte classant



En cas de transfert avant 5 jours, il y a très peu d'actes classants.



### Bilan empirique

- Règles expertes :
  - Hautement valides
  - Généralement peu contributives (~0% des gains) car chaque situation est plus complexe que ce que la règle ne prend en compte
  - => plus pour améliorer la qualité mesurable (Datim, Ovalide) que pour augmenter les recettes
- Règles Machine Learning
  - Très impressionnantes : suggèrent des situations non anticipées
  - En pratique cependant, pas indispensables : ~20% des gains
- Règles maladies chroniques
  - Valides et très efficaces (~80% des gains)
  - Importance de la qualité des mappings
  - VPP < 100% cependant : vérifier que la maladie est toujours active et qu'elle est effectivement prise en charge
- Métarègles :
  - Simples et efficaces
  - En théorie pas nécessaires en Machine Learning, mais en pratique utiles car traitent des exceptions relativement rares

# Software developed for automated data quality control (2008)

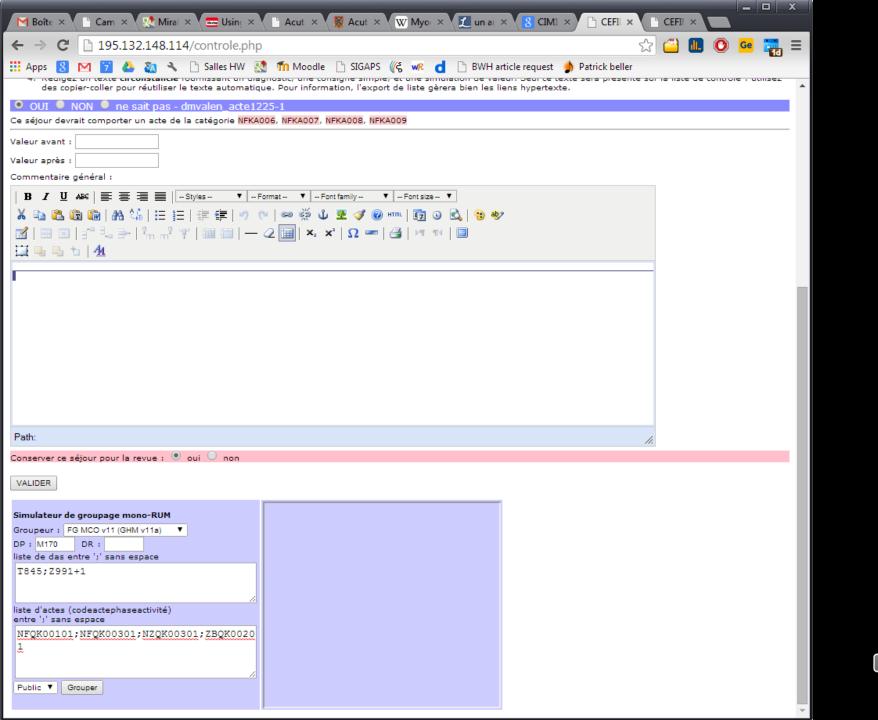
Idea and Software design:

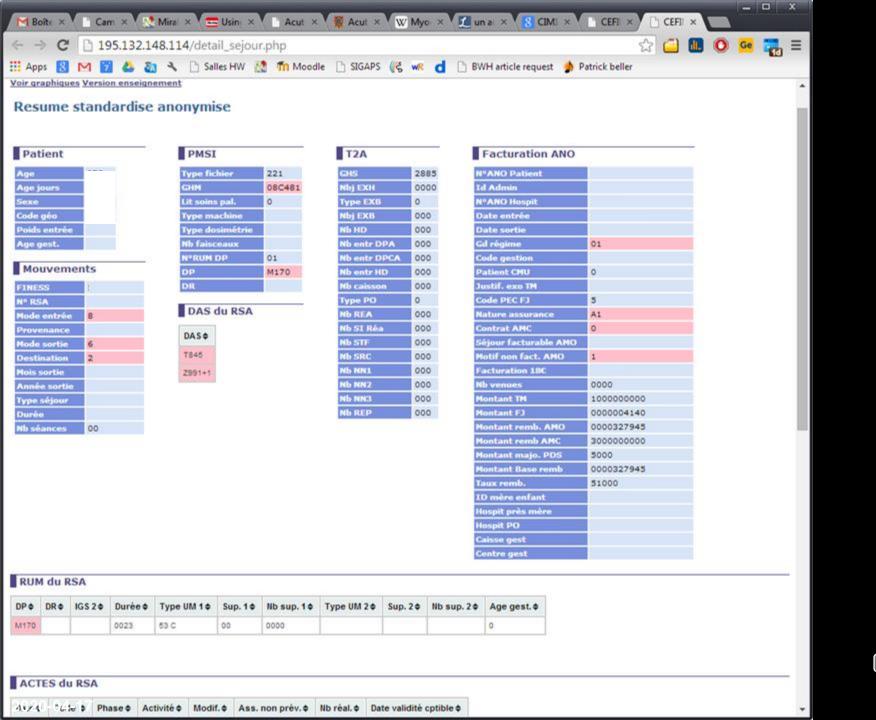
E Chazard, M Genty & G Ficheur

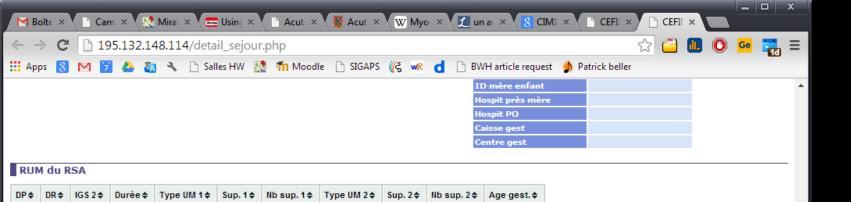
Data analysis & software development:

**E** Chazard









ACTES du RSA

0023

53 C

M170

Date <b>♦</b>	Code \$	Phase \$	Activité <b></b>	Modif. <b>♦</b>	Ass. non prév.≑	Nb réal.≑	Date validité cptible \$
001	NEKA020	0	4			1	1
001	NEKA020	0	1			1	1
001	NFQK001	0	1	BZ	1	1	1
000	NFQK003	0	1	z	2	2	1
000	NZQK003	0	1	Z	1	1	1
000	ZBQK002	0	1	Z	2	1	1

00

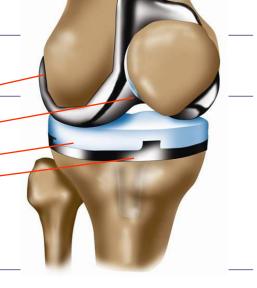
0000

#### Molecules onereuses

Pas d'enregistrement

#### Dispositifs medicaux implantables

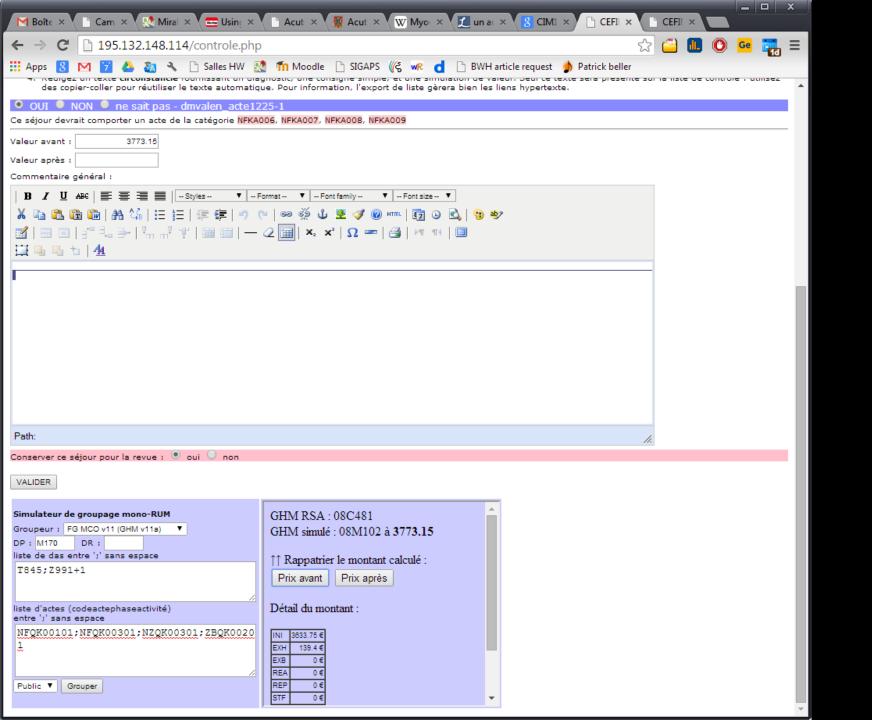
Dispositiis ii	neuicaux iiii	лансарі	es				
Année séjour¢	Mois séjour¢	Code <b></b>	Nombre \$	Montant <b></b>	Mois <del>pos</del> e <b>\$</b>	Année pose \$	Date réal.◆
2014	02	3127942	0000000001	0001292430	02	2014	0001
2014	02	3152319	0000000001	0000732920	02	2014	0001
2014	02	3157570	0090000001	0000133240	02	2014	0001
2014	02	3163659	020000001	0000039560	02	2014	0001
2014	02	3181870	0.00000001	0000359740	02	2014	0001

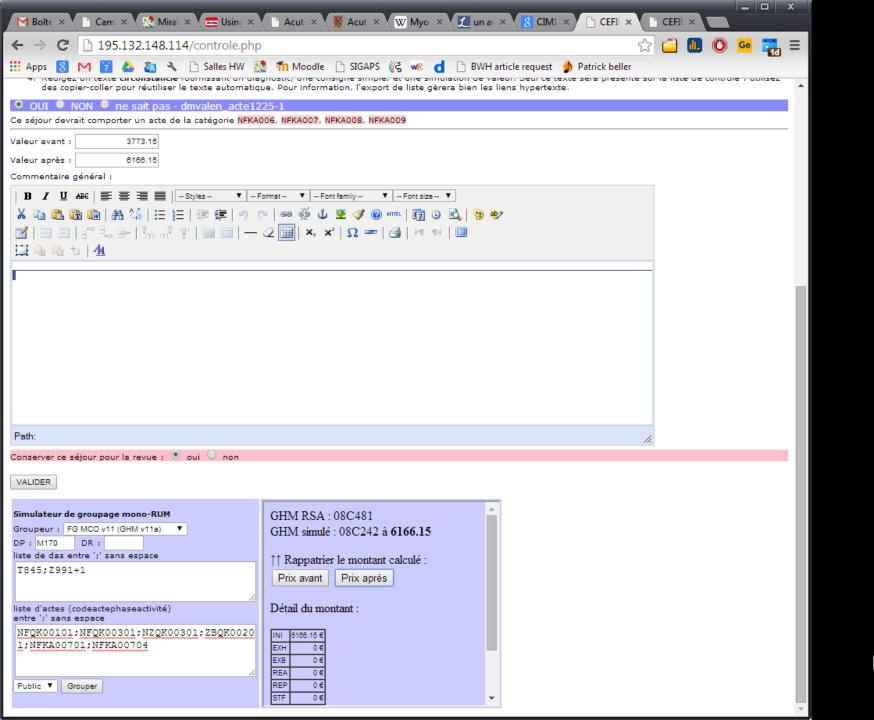


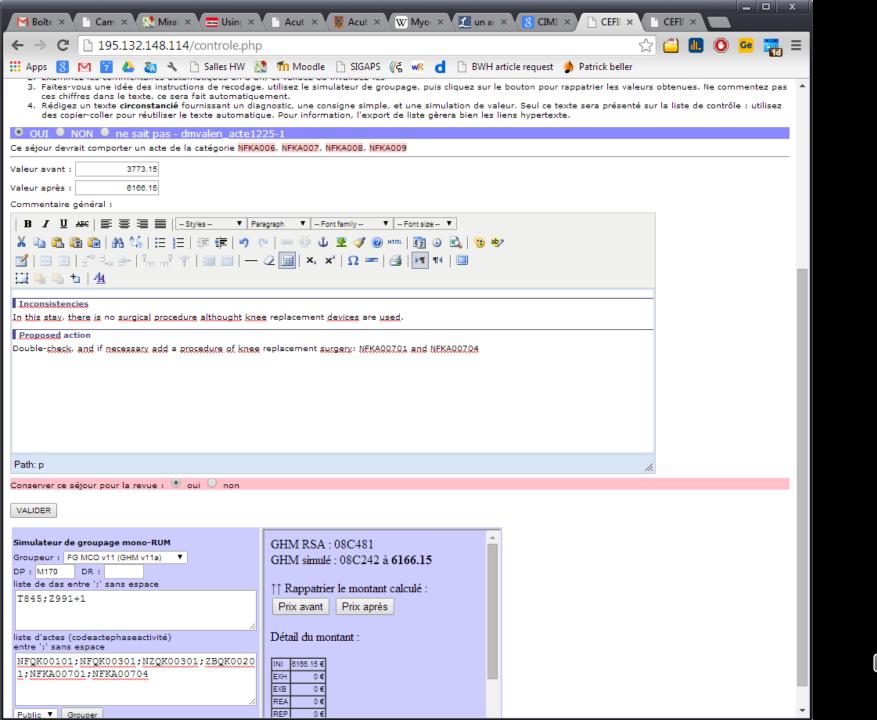


Pas d'enregistrement











### Retour d'expérience

- Gains variables selon la qualité initiale du codage : mauvaise qualité initiale => gains élevés
- Exemples donnés ici : pour une mission courte, 2 semaines max
- Gains habituels en hospitalisation complète MCO
  - Proposition de recodage sur 1,5% des séjours hors séances
  - Gain moyen de 1200€ par séjour
  - Exemple:
    - 100 000 séjours contrôles
    - 1 500 séjours valorisants
    - gain habituellement observé après validation DIM de 1,35 millions €
- Gains habituels en soins externes MCO
  - Gains entre 20 000€ et 390 000€ pour 100 000 épisodes externes
  - Très variable selon la qualité des contrôles déjà mis en place à la saisir



### Merci de votre attention!

emmanuel.Chazard@univ-lille.fr