

Réutilisation de données hospitalières et intelligence artificielle : des données à l'intervention de santé, un chemin cahoteux

Pr Emmanuel Chazard

Université de Lille, CERIM, ULR 2694

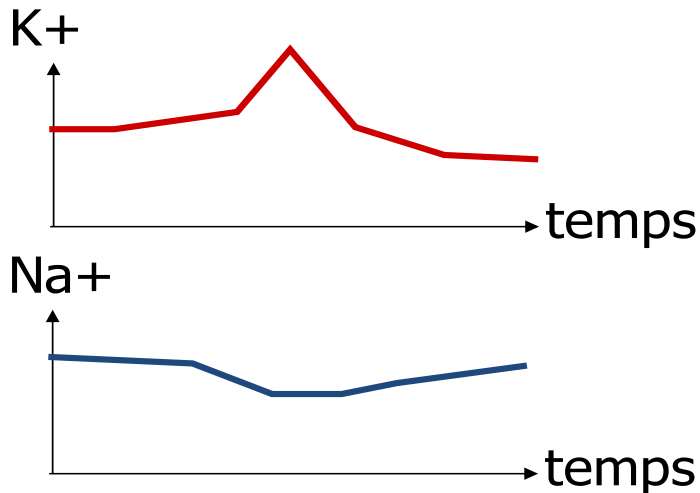
CHU de Lille, pôle de Santé Publique

Structuration de cette présentation

- Cas pratique :
 - Données : séjours hospitaliers (réutilisation de données)
 - Données du PMSI
 - Résultats d'analyses de biologie médicale
 - Médicaments administrés
 - Objectifs :
 - Identifier un événement péjoratif (ex : surdosage en anticoagulants, hyperkaliémie)
 - Automatiquement identifier les facteurs de risque
- Au fil des étapes du cas pratique, 18 leçons apprises...

Présentation des données fonctionnelles disponibles

- Données fonctionnelles = dont la valeur dépend du temps
- Ex 1 : Résultats d'analyses de biologie médicale
« la biologie »



Id Patient	Type	Date	Valeur
123	K+	13/02/2011	4.3
123	K+	15/02/2011	4.9
123	K+	16/02/2011	5.3
...			
123	NA+	13/02/2011	140
123	NA+	15/02/2011	138
123	NA+	16/02/2011	135

- Ex 2 : Médicaments administrés au patients
« les médicaments »

En moyenne
20 valeurs
par séjour

En moyenne
100 valeurs
par séjour

Chargement des données dans l'entrepôt

Leçon 1 :
terminologies

- Problème : pas de normalisation des libellés, pas de terminologie
 - Théorie : utiliser LOINC « 6298-4 »
 - Pratique :
 - libellés locaux « K+ », « potassium sang », « kaliémie », « K1 »...
 - Non interopérables entre établissements
 - Variables dans le temps
 - Variables entre automates

Id Patient	Type	Date	Valeur
123	K+	13/02/2011	4.3
123	K+	15/02/2011	4.9
123	K+	16/02/2011	5.3

Chargement des données dans l'entrepôt

Leçon 2 : qualité orientée par l'utilisation

- Problème : les unités sont parfois fausses

■ Ex :

HEMATIES	4. 580 / mm3	N: 4,0-5,3
Hémoglobine	14,1 g/dl	N: 12-16
Hématocrite	42 %	N: 37-46

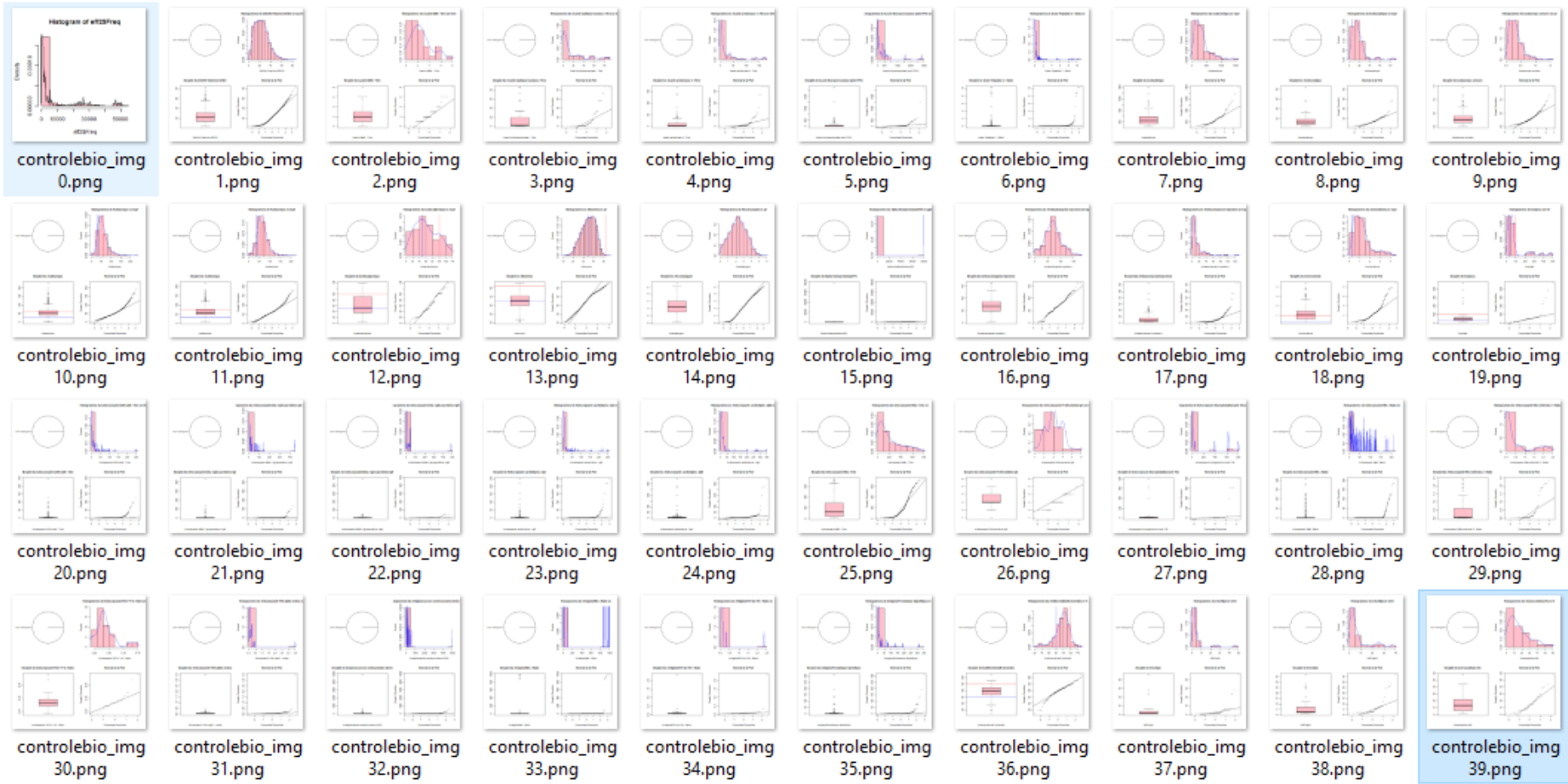
- Problème : les données sont « prêtes à être imprimées » mais peu normalisées

■ Ex :

Id Patient	Type	Date	Valeur	Unité
123	Glycémie	13/02/2011	5,5	mmol/l
123	Soit en g/l	13/02/2011	1,00	g/l

Contrôle qualité des données

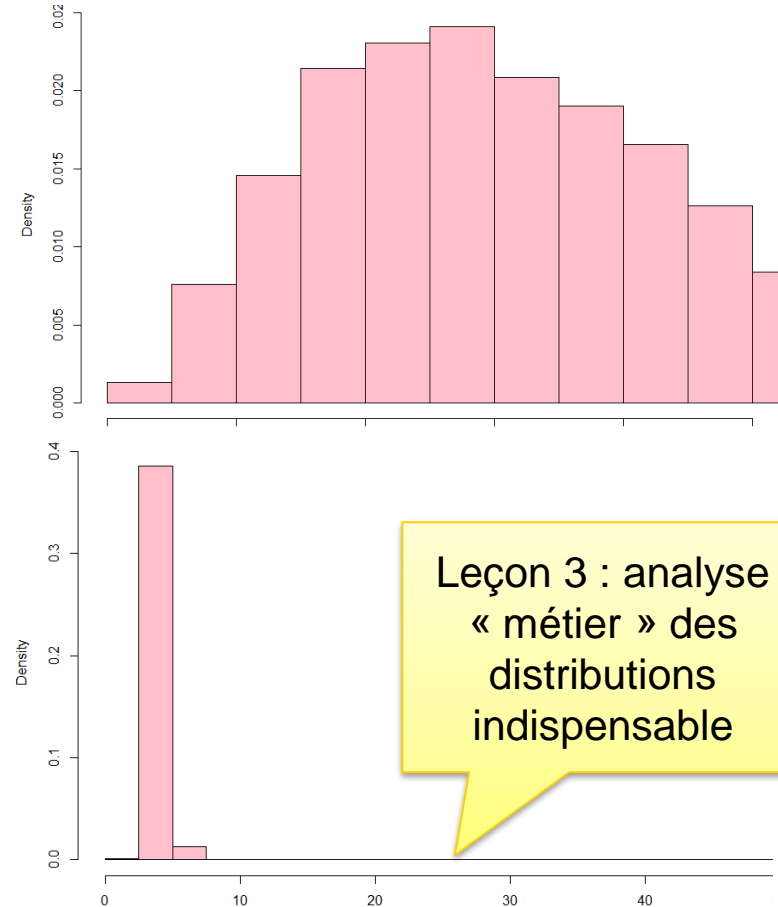
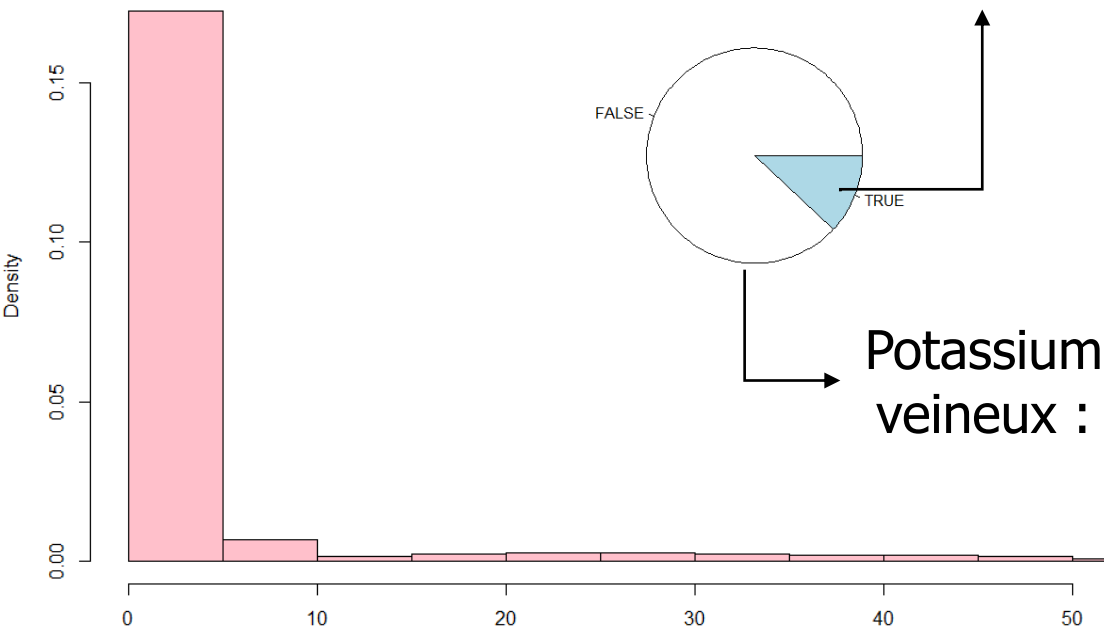
ex : 474 paramètres de biologie médicale



Contrôle qualité des données

ex : 474 paramètres de biologie médicale

Paramètre nommé
« Potassium »



Leçon 3 : analyse
« métier » des
distributions
indispensable

Comment traiter ces paramètres ?

- Traitement direct impossible :
 - Trop de paramètres
 - Trop de valeurs (une par point de la courbe, ~100 par séjour)
- On aimerait « une ligne par patient »
- Nécessité de grandement simplifier les données

id_stay	date	label	value	unit
607755990	3	PE /mm3	208	/mm3
607755990	3	P. éosinophiles	3	%
607755990	3	PN (/mm3)	4913	/mm3
607755990	3	P. neutrophiles	71	%
607755990	3	I.N.R.	1.15	
607755990	3	Taux de prothrombine	82	%
607755990	3	Rapport P/T	1.25	
607755990	3	T.C.A. patient	45	sec
607755990	3	T.C.A. témoin	36	sec
607755990	3	Réserve alcaline	25	mEq/l
607755990	3	Protides totaux	57	g/L
607755990	3	PLAQUETTES x1 000	177	/mm3
607755990	3	Sodium	136	mEq/l
607755990	3	Potassium	4.3	mEq/l
607755990	3	Créatinine	11	mg/L
607755990	3	Chlore	98	mEq/l

Leçon 4 : impossible qu'un clinicien requête directement de telles données

Administrative data

88 years old woman

Diagnoses

- I10 Arterial hypertension
- Z8671 Personal history of myocardial ischemia
- I620 Non-traumatic subdural hemorrhage

Medical procedures

- ABJA002 Drainage of an acute subdural hemorrhage, by craniotomy
- FELF001 Transfusion

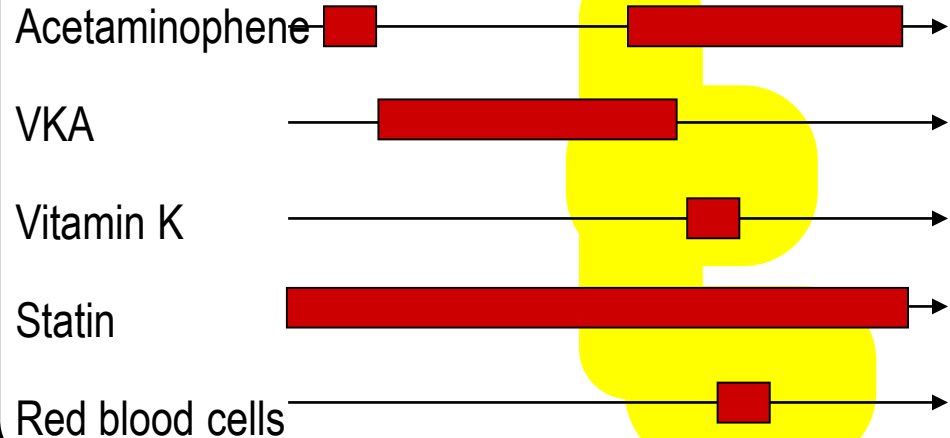
Free-text reports

Discharge letter

Surgical report

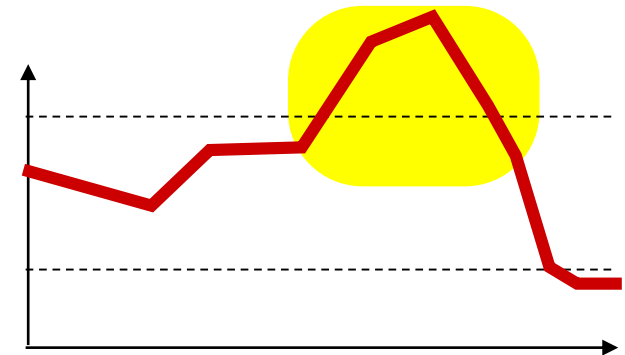
Leçon 5 : impossible qu'un informaticien devine ce qu'il faut faire

Drugs

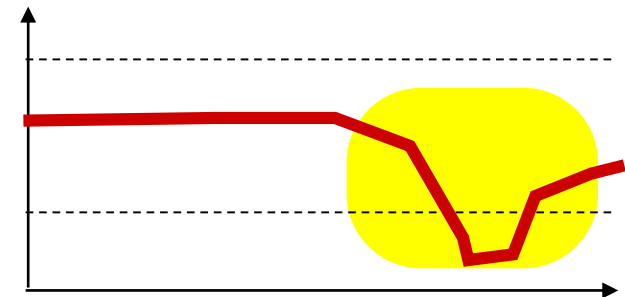


Laboratory results

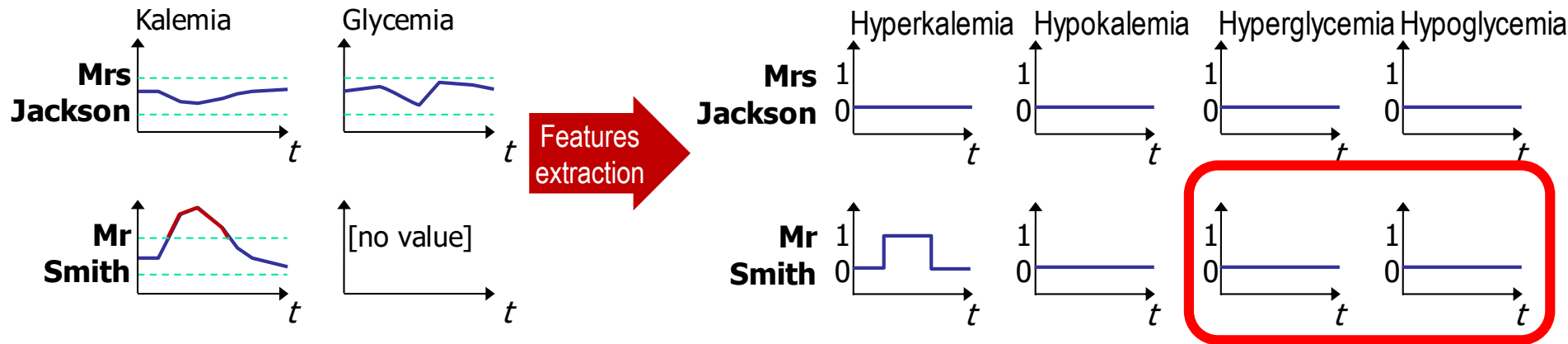
INR



Hemoglobin



Extraction de caractéristiques dans les résultats de biologie médicale



Example of missing data handling

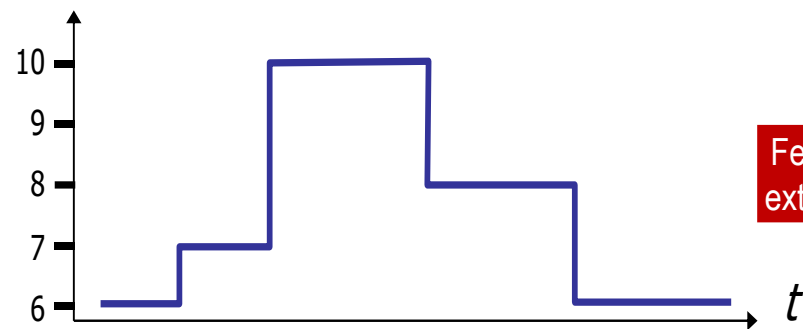
■ Formally:

- Example of 2 patients, 2 parameters measured 5 times
- Before: 1 table with 2 lines + 1 table with 10 lines
- After: 1 table with 2 lines

Leçon 6 : traitement de données nécessitant une algorithmique et une connaissance métier

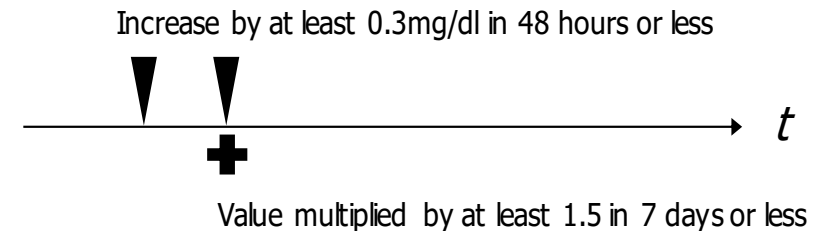
Extraction de caractéristiques dans les résultats de biologie médicale : exemple plus complexe

Creatinine :



Features
extraction

Acute renal failure (KDIGO) :



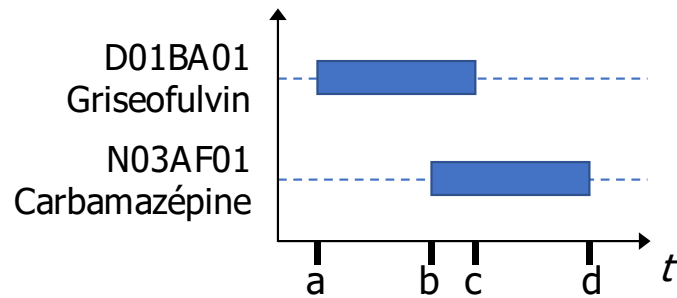
■ Formally:

- Example of 1 patients, 6 measurements of creatinine
- Before: 1 table with 1 line + 1 table with 6 lines
- After: 1 table with 1 line

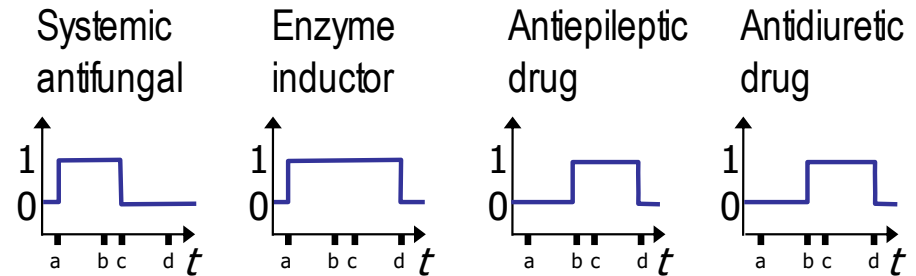
Leçon 7 :
nécessité
néanmoins de
partager certaines
fonctions pour
critères
internationaux

Extraction de caractéristiques dans les médicaments administrés

Drugs administered to Mrs Jackson



Features
extraction



- Formally:
 - Example of 1 patient, 2 administered drugs
 - Before: 1 table with 1 line + 1 table with 2 lines
 - After: 1 table with 1 line

Leçon 8 :
extraction de
caractéristiques
dépend
fortement de
l'étude, non
générique

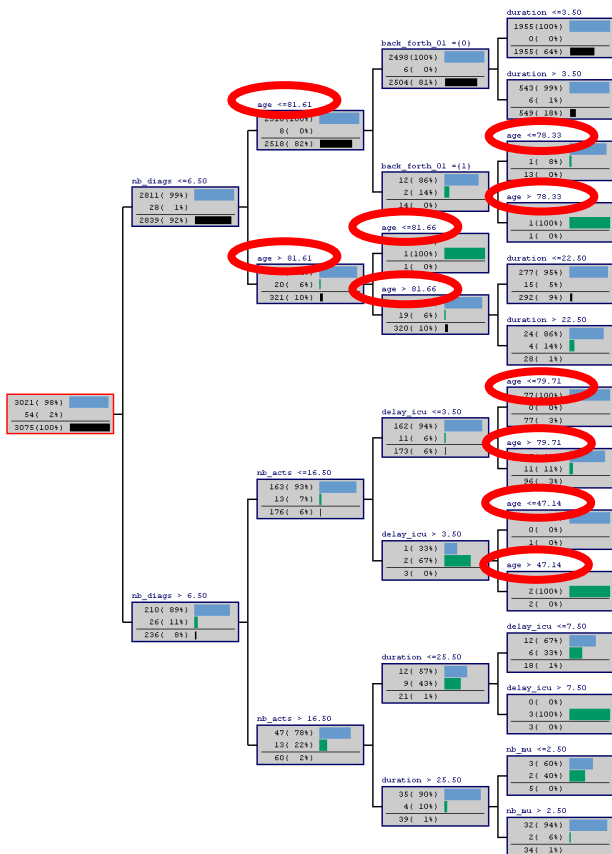
Recherche d'associations à risque de décès

Leçon 9 : bien connaître les méthodes statistiques

Variables explicatives : ~20 variables démo/administratives, ~500 paramètres de biologie, ~500 médicaments. L'âge revient sans cesse...

Leçon 10 : optimiser les données pour répondre à la question

Leçon 11 : spécialiser les variables à expliquer : les patients sont là parce qu'ils sont malades !

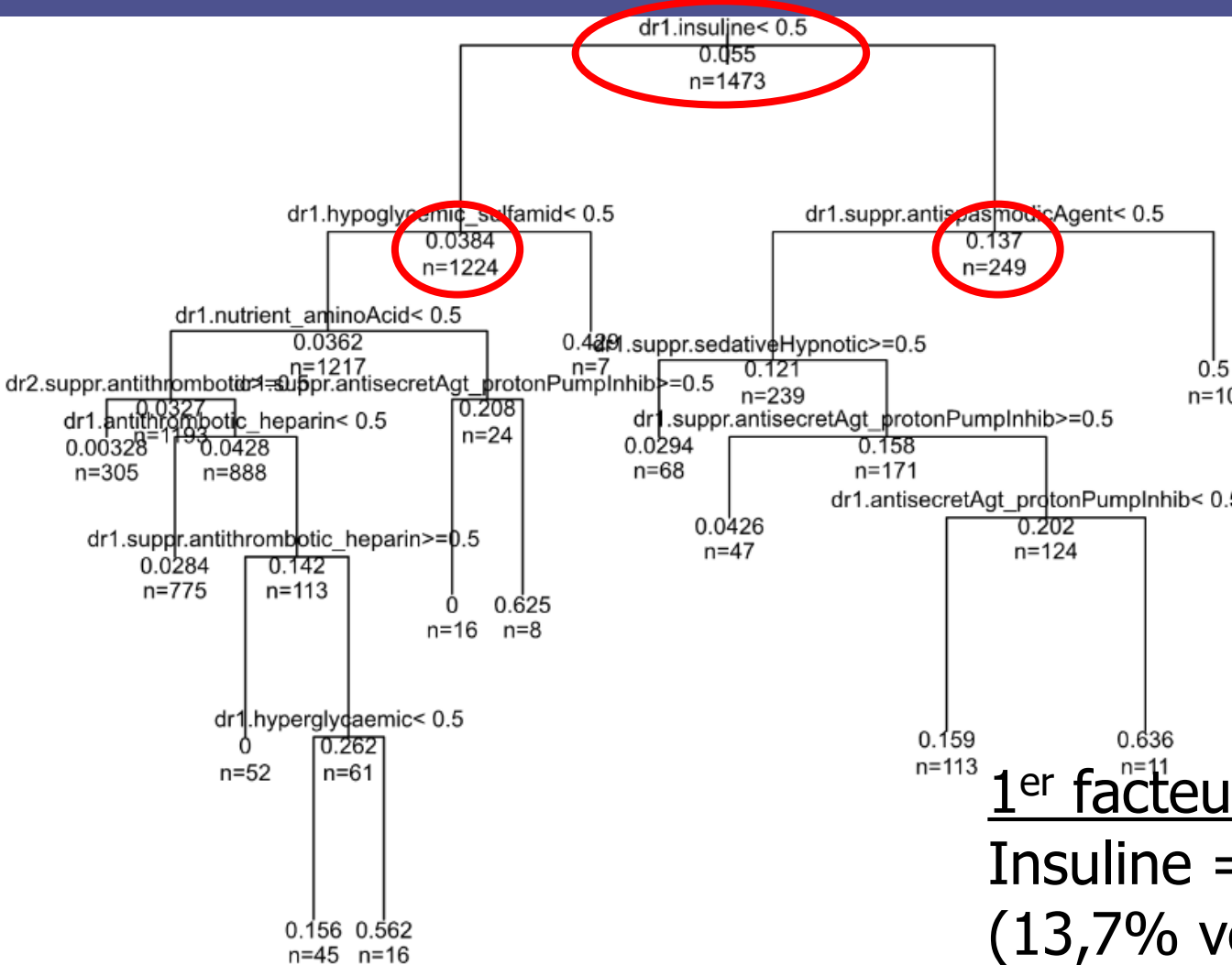


Base de connaissances : 15 règles

N°	N°SEGMENT	Condition d'appartenance	Support	Conclusion	AE_1 (C7=0)	AE_2 (C7=1)
1	10	SI age <=81.66 et age > 81.61 et nb_diags <=6.50	0.0000	AE_1 (C7=0)	0.0000	0.0000
2	16	SI duration <=3.50 et back_forth_01 = {0} et age <=81.61 et nb_diags <=6.50	0.0000	AE_1 (C7=0)	0.0000	0.0000
3	17	SI duration > 3.50 et back_forth_01 = {0} et age <=81.61 et nb_diags <=6.50	0.0000	AE_1 (C7=0)	0.0000	0.0000
4	18	SI age <=78.33 et back_forth_01 = {1} et age <=81.61 et nb_diags <=6.50	0.0000	AE_1 (C7=0)	0.0000	0.0000
5	19	SI age > 78.33 et back_forth_01 = {1} et age <=81.61 et nb_diags <=6.50	0.0000	AE_1 (C7=0)	0.0000	0.0000
6	20	SI duration <=22.50 et age > 81.66 et age > 81.61 et nb_diags <=6.50	0.0000	AE_1 (C7=0)	0.0000	0.0000
7	21	SI duration > 22.50 et age > 81.66 et age > 81.61 et nb_diags <=6.50	0.0000	AE_1 (C7=0)	0.0000	0.0000
8	22	SI age <=79.71 et delay_icu <=3.50 et nb_acts <=16.50 et nb_diags <=6.50	0.0000	AE_1 (C7=0)	0.0000	0.0000
9	23	SI age > 79.71 et delay_icu <=3.50 et nb_acts <=16.50 et nb_diags > 6.50	0.0000	AE_1 (C7=0)	0.0000	0.0000
10	24	SI age <=47.14 et delay_icu > 3.50 et nb_acts <=16.50 et nb_diags > 6.50	0.0000	AE_1 (C7=0)	0.0000	0.0000
11	25	SI age > 47.14 et delay_icu > 3.50 et nb_acts <=16.50 et nb_diags > 6.50	0.0000	AE_1 (C7=0)	0.0000	0.0000
12	26	SI delay_icu <=7.50 et duration <=25.50 et nb_acts > 16.50 et nb_diags > 6.50	0.0000	AE_1 (C7=0)	0.0000	0.0000
13	27	SI delay_icu > 7.50 et duration <=25.50 et nb_acts > 16.50 et nb_diags > 6.50	0.0010	AE_2 (C7=1)	0.0000	1.0000
14	28	SI nb_mu <=2.50 et duration > 25.50 et nb_acts > 16.50 et nb_diags > 6.50	0.0016	AE_1 (C7=0)	0.6000	0.4000
15	29	SI nb_mu > 2.50 et duration > 25.50 et nb_acts > 16.50 et nb_diags > 6.50	0.0111	AE_1 (C7=0)	0.9412	0.0588

Date de création : 02/05/2008 16:39:16

Recherche d'associations : hyperglycémie

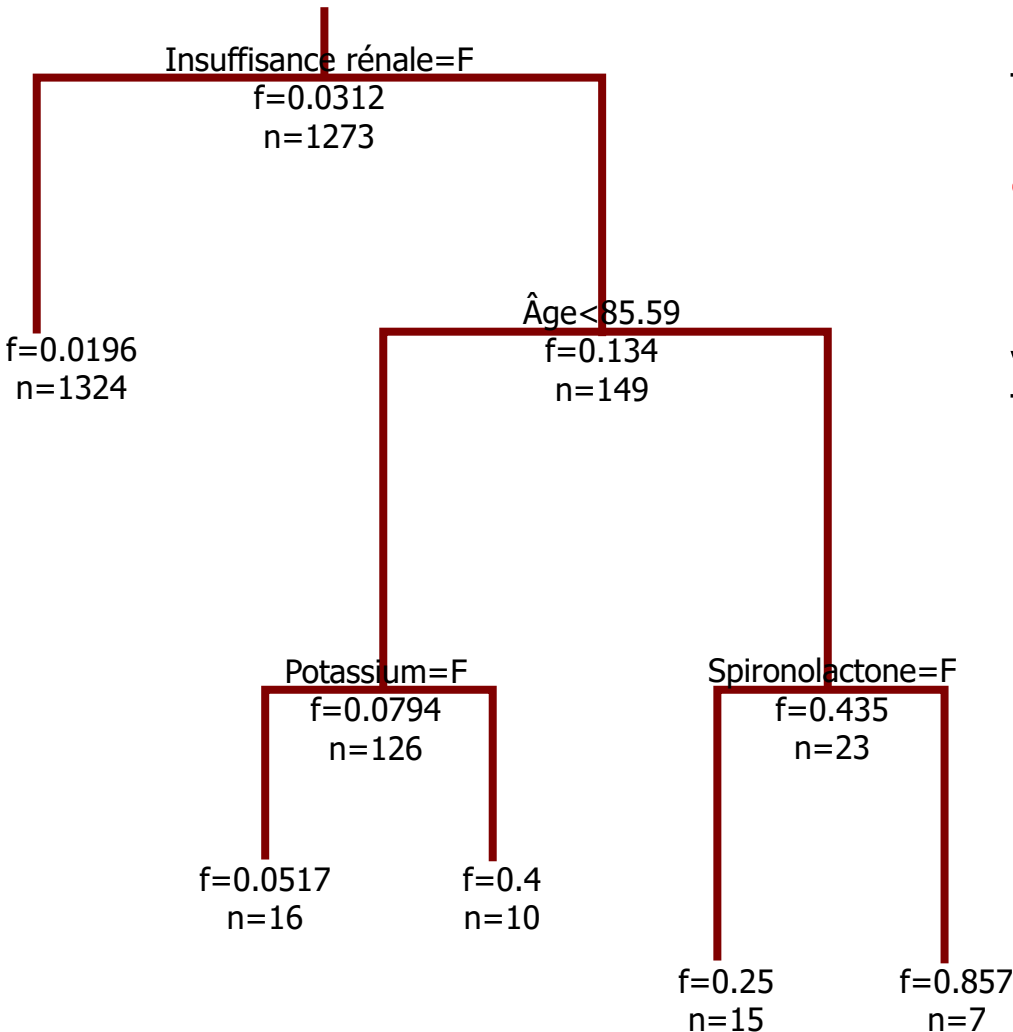


Leçon 12 : savoir interpréter les données et filtrer les résultats

Leçon 13 : donc machine learning en « black box » pas toujours utilisable

1^{er} facteur de risque ☹ :
 Insuline => Hyperglycémie ?
 (13,7% versus 3,8%)

Recherche d'associations : hyperkaliémie



Règle le plus à droite ☺ :
insuffisance rénale & âge>85
& spironolactone
=> 6 hyperK⁺ sur 7 exposés

Validation :

- Technique ☺
- Statistique ☺
- Pharmacologique ☺
- Revue des cas
(lecture des dossiers) ☹ :

 - 3 EIM confirmés
 - 2 lyses tumorales
 - 1 fin de vie

Leçon 14 : toujours revenir
aux cas cliniques... moins
évidents que prévu

Contextualisation par service

Rule: vitamin K antagonist & amoxicilline&clav.ac. & age ≥ 70 → appearance of high INR

Department	Confidence (PPV)	Support (frequency)	Median delay	Relative risk	Fisher's test P value
X all departments	10/57=17.5%	10/5322=1.9‰	6.5j	13.38	0
X surgery	0/3=0%	0/1055=0‰		0	1
X gyneco-obstetrics	<i>No stay matches the conditions</i>				
X medicine A	1/11=9.1%	1/1236=0.8‰	7j	5.06	0.1874
X medicine B	3/17=17.7%	3/966=3.1‰	3j	9.31	0.005
X pneumology	5/28=17.9%	5/818=6.1‰	11j	6.41	0.0016
Y all departments	1/10=10%	1/11923=0.1‰	6j	33.09	0.0306
Y apoplexy	<i>No stay matches the conditions</i>				
Y cardio & endocrinology	1/2=50%	1/1967=0.5‰	6j	51.71	0.0202
Y geriatrics	0/2=0%	0/493=0‰		0	1
Y gynecology	<i>No stay matches the conditions</i>				
Y intensive care unit	<i>No stay matches the conditions</i>				
Y internal medicine	0/5=0%	0/1514=0‰		0	1
Y obstetrics	<i>No stay matches the conditions</i>				
Y orthopedics	<i>No stay matches the conditions</i>				
Y rheumatology	<i>No stay matches the conditions</i>				
Y urology	<i>No stay matches the conditions</i>				
Z all departments	0/1=0%	0/1022=0‰			
W all departments	0/8=0%	0/7685=0‰			

Leçon 15 : le risque n'est pas le même partout (site = proxy pour les variables latentes...)

Proposition d'une approche contextualisée

- Ex : AVK & IPP → risqué hémorragique
- Implémentation habituelle :

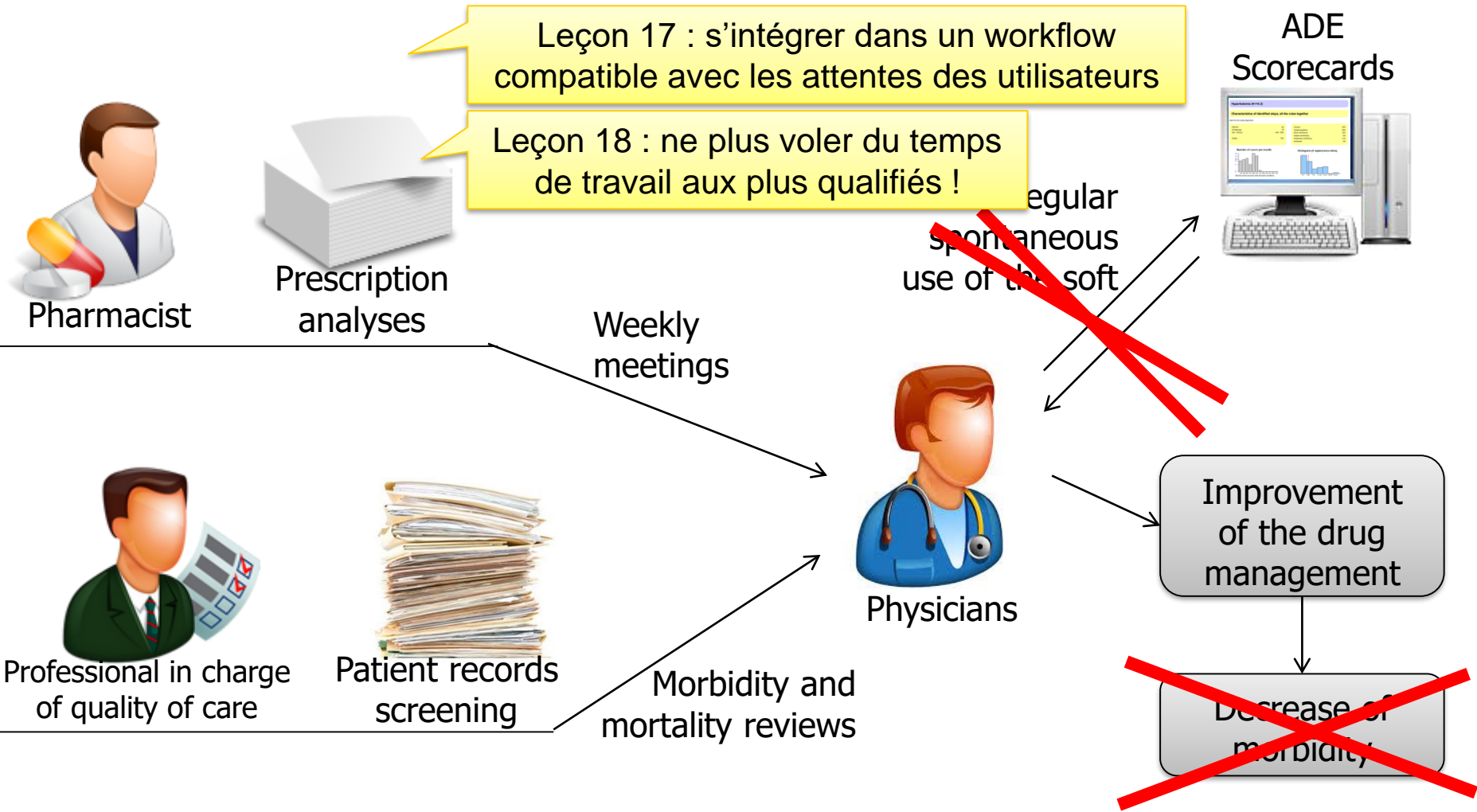
Medical unit A	Medical unit B	Medical unit C
VKA & PPI → interruptive alert		

- Proposition de contextualisation :

Medical unit A	Medical unit B	Medical unit C
Empirical probability=10% VKA&PPI→ interruptive alert	Empirical probability=0.01% VKA&PPI→ <i>silent or non-interruptive alert</i>	Unseen circumstances VKA&PPI→ interruptive alert

Leçon 16 : tout faire pour réduire le bruit (over-alerting) et économiser les nerfs des utilisateurs

Conception d'un outil d'aide à la prescription basé sur ces résultats



Au bilan

- Chemin semé d'embûches
 - 18 leçons apprises, et bien d'autres !
 - Rassurez-vous : nous sommes arrivés à nos fins !
- En particulier, indispensables en même temps :
 - Informatique, algorithmique
 - Statistique
 - Forte connaissance médicale, tant sur les données sources que la finalité du traitement
 - Liant : talent pour traduire le problème médical en termes compréhensibles et acceptables pour les informaticiens et statisticiens
 - Artisanat : impact majeur de l'extraction de caractéristiques
 - Facteurs humains et économie : intégration des éventuels produits dans un workflow acceptable et économiquement efficient

Merci de votre attention

emmanuel.chazard@univ-lille.fr