

## **Les « ADE Scorecards » : un outil de détection par data mining des effets indésirables liés aux médicaments dans les dossiers médicaux (projet PSIP)**

**Emmanuel Chazard<sup>1</sup>, Adrian Baceanu<sup>2</sup>,  
Grégoire Ficheur<sup>1</sup>, Romaric Marcilly<sup>3</sup> et Régis Beuscart<sup>1</sup>**

<sup>1</sup>*Service d'information et des archives médicales, CHRU de Lille, EA2694, Lille, France*

<sup>2</sup>*IDEEA Advertising, Bucarest, Roumanie*

<sup>3</sup>*Evalab INSERM CIC-IT, Univ Lille Nord de France, CHRU Lille, EA 2694, Lille, France*

### **Abstract**

*The automated detection of Adverse Drug Events (ADE) is an important issue in medical informatics. The objective of this work is to automatically detect ADEs and to present the results to physicians. 90,000 stays are extracted from the EHR of 5 French and Danish hospitals and loaded into a common repository, using a common data model. Then data mining procedures such as decision trees are used in order to get ADE detection rules that are filtered and validated by an expert committee. The procedure enables to produce 236 ADE detection rules that are able to detect 27 different kinds of outcomes. Contextualized statistics are computed for every rule in every medical department separately. The physicians of the medical departments are provided with that information by means of a web-based tool named "ADE Scorecards". The tool is presented in the article through a use case and several screenshots. Based on a list of rules and a repository of stays, it allows for displaying the important rules, the related statistics, and the complete information about the suspicious cases. The knowledge is contextualized, i.e. it depends on the medical department. The tool is deployed in a French hospital and is currently being evaluated through a prospective impact assessment.*

### **Keywords**

Adverse Drug Events; Data Mining; Decision Trees; Electronic Health Records

## **1 Introduction**

Les effets indésirables liés aux médicaments (EIM) sont un enjeu de santé publique. Ils seraient responsables chaque année de 15 000 décès en France, et 98 000 aux USA [1]. Les méthodes les plus usuelles de détection des cas d'EIM reposent sur la déclaration volontaire, or il est prouvé que cette déclaration concerne moins de 5% des EIM [2-3], et la revue de dossiers, efficace mais très chronophage, et ce d'autant plus que les EIM sont relativement rares. Des méthodes de *natural language processing* (NLP) sont appliquées aux courriers de sortie, ce qui suppose néanmoins que l'EIM soit cité dans le courrier [4]. Enfin, des méthodes de data mining (fouille statistique de données) peuvent être appliquées aux déclarations d'EIM [5-6], ce qui suppose que les EIM soient déclarés, ce qui est rarement le cas.

La connaissance sur les EIM est essentiellement académique. Les prescripteurs trouvent habituellement les informations nécessaires dans les résumés des caractéristiques du produit (RCP) disponibles pour chaque médicament, et tenues à jour en France par l'AFSSAPS. Cependant, ces informations sont en grande quantité et submergent le lecteur (ex : 3 300 mots pour le Previscan). L'information est classée par importance théorique, or les EIM surviennent avec des probabilités différentes et souvent inattendues, en fonction des connaissances des prescripteurs et de leurs pratiques de surveillance du risque [7]. Le projet PSIP (Patient Safety through Intelligent Procedures in medication) a pour objectifs de générer de la connaissance sur les EIM notamment par data mining, et de mettre en place des outils novateurs de prévention de ces EIM. Le travail présenté ici consiste à collecter des dossiers patient électroniques et à les fouiller à l'aide de méthodes de data mining pour détecter automatiquement des cas passés d'EIM, et identifier les circonstances en relation avec ces EIM. Enfin, l'objectif est de présenter aux médecins de chaque service cette connaissance de manière intelligible, ainsi que les cas d'EIM détectés par le système. La présentation de connaissance sera contextualisée en fonction du service.

## 2 Matériel et méthodes

### 2.1 Constitution d'une base de données centralisée

Une base de données centralisée est mise en place : elle est destinée à regrouper les séjours des différents hôpitaux. Un modèle de donnée est tout d'abord mis au point. Il représente des informations administratives et démographiques (identifiants, parcours, âge, sexe...), les diagnostics du patient codés en CIM10, les actes thérapeutiques ou diagnostiques codés dans la nomenclature nationale, les médicaments administrés codés en ATC, les résultats de biologie codés en IUPAC, et des éléments en texte libre (ex : courrier de sortie). Lorsque l'hôpital ne dispose pas d'un circuit du médicament informatisé, les codes ATC correspondant aux médicaments prescrits sont automatiquement extraits des courriers de sortie à l'aide de l'outil d'extraction F-MTI [8].

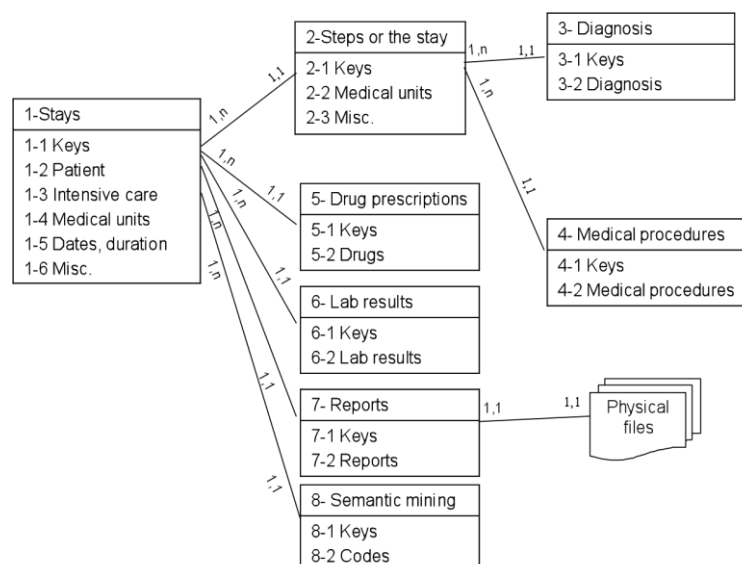


Figure 1 : Représentation simplifiée du modèle de données

Le modèle de données est partiellement normalisé afin d'en accroître la performance et de faciliter l'extraction de données [9]. En outre, il semble compatible avec tous les systèmes d'information qui recueillent ces types d'information. Ce modèle a été publié [10], la

figure 1 en présente une vue simplifiée.

Les données sont extraites des dossiers patients électroniques de l'hôpital général de Denain (Fr, 39 010 séjours), l'hôpital de Nordsjaelland (Da, 23 067 séjours), l'hôpital de Frederiksberg (Da, 21 331 séjours), le CHRU de Lille (F, 7 711 séjours) et le CHU de Rouen (F, 1 367 séjours). Il s'agit de données collectées en routine, il n'y a aucun ajout d'information spécifique. Au total, nous disposons ainsi de 92 686 séjours.

## 2.2 Mise au point d'une liste de manifestations possible d'EIM

Dans un premier temps, il nous faut identifier les événements nuisibles qui pourraient être le fait d'EIM. Les Résumés des Caractéristiques du Produit (RCP) constituent l'information officielle disponible sur tous les médicaments mis sur le marché. Ces RCP sont maintenues par l'Agence Française de Sécurité Sanitaire des Produits de Santé (AFSSAPS) et édités notamment par Vidal SA. Chaque élément de connaissance est structuré sous forme de règle, dans laquelle des causes (par exemple héparine et insuffisance rénale) peuvent conduire à un effet (par exemple hémorragie).

Les règles issues des RCP permettent de lister 228 manifestations possibles des EIM, cliniques et para cliniques. Dans les données dont nous disposons, 83 de ces effets sont traçables, à travers 56 variables différentes. La liste initiale contient effectivement plusieurs synonymes, comme par exemple « hypoxémie », « hypoxémie artérielle » et « PaO<sub>2</sub> < 60 mm Hg au repos ».

## 2.3 Induction de règles de détection des EIM par fouille statistique des données (data mining)

### 2.3.1 Agrégation des données

Afin de préparer la fouille de données et le calcul d'indicateurs statistiques, les données sont agrégées sous la forme simplifiée d'événements. Un événement peut se produire (il vaut alors 1) ou ne pas se produire (il vaut alors 0). S'il se produit, il peut être caractérisé par une date de début et une date de fin.

La figure 2 montre l'exemple simple d'un séjour pour lequel une hyperkaliémie est détectée. L'événement « hyperkaliémie » est créé entre j2 et j6. L'événement hypokaliémie vaut 0 car aucune valeur trop basse n'est observée.

L'agrégation des données est réalisée pour toutes les données disponibles : les diagnostics (48 événements correspondant à des maladies chroniques), les variables administratives et démographiques (15 événements), les médicaments administrés (250 événements), et les résultats de biologie (35 événements). Certains de ces événements sont des effets potentiels des EIM, d'autres sont des causes potentielles d'EIM, et un nombre important peuvent être l'un comme l'autre. Au total, 588 variables pourront être utilisées pour expliquer 56 effets différents.

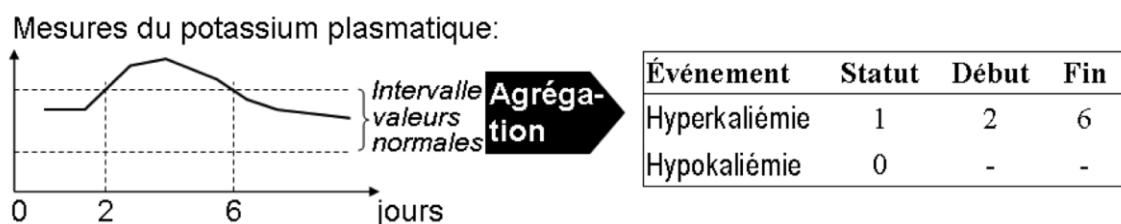


Figure 2 : Exemple d'agrégation de données de biologie : séjour avec hyperkaliémie

### 2.3.2 Induction supervisée de règles de détection des EIM

La détection des cas d'EIM est réalisée à l'aide de règles. Une règle associe un ensemble de conditions booléennes qui, en cas de réalisation, permettent de prédire la réalisation d'un effet, selon la forme  $SI ( condition\_1 \& \dots \& condition\_k ) ALORS\ effet$ . Des arbres de décision suivant la méthode CART sont réalisés, grâce au package RPART du logiciel de programmation en statistiques R [11-13], ainsi que des règles d'association [14]. Ces deux méthodes permettent de générer des règles suivant le formalisme ci-dessus. Elles sont invoquées pour chaque type d'effet tracé, et pour chaque service hospitalier disponible. Des milliers de règles sont générées automatiquement. Lors de l'induction de règles permettant de prédire un effet donné, les données sont modifiées à la volée afin de maintenir une cohérence temporelle dans les règles produites : au moment de prédire un effet donné, seuls les événements antérieurs à cet effet et encore actifs au moment de survenue de cet effet sont pris en compte. Un exemple d'arbre de décision est décrit dans la section *Résultats* de cet article.

### 2.3.3 Filtrage et formatage des règles

Les règles produites par les arbres de décision et les règles d'association sont filtrées automatiquement. Seules les règles en rapport avec l'usage du médicament sont conservées, *i.e.* celles qui contiennent l'administration d'un médicament, la suppression d'un médicament, ou un événement biologique en rapport implicite avec un médicament (comme par exemple un dosage plasmatique de médicament). Les règles sont ensuite soumises à un comité d'experts comprenant des médecins, pharmacologues et pharmaciens. Ce comité ne valide que les règles pertinentes par rapport à la connaissance académique. L'induction de règles tient parfois compte de la connaissance académique. Par exemple, les règles prédisant la survenue d'une hyperkaliémie sont analysées séparément chez les insuffisants rénaux et chez les non insuffisants rénaux. Les règles ainsi validées sont décrites sous forme d'objets XML dans une base de connaissance. A la demande des experts, certaines règles académiques sont ajoutées à la base de connaissance. Cela concerne en particulier les événements absents ou quasi-absents des données étudiées : le data mining ne peut générer de règle pertinente en-deçà de cinq cas.

### 2.3.4 Calcul automatisé et multi-site de statistiques caractérisant les règles

Chaque règle de la base de connaissance est automatiquement évaluée dans chaque service. Cette évaluation automatisée permet d'ajouter des informations statistiques contextualisées, c'est-à-dire susceptibles de changer d'un service à un autre. Sachant qu'une règle est un ensemble de conditions  $(C_1 \& \dots \& C_k)$  qui mènent à un effet  $(E)$ , les statistiques suivantes sont calculées en tenant compte des contraintes temporelles :

- $support = P(E \cap C_1 \cap \dots \cap C_k)$
- $confiance = P(E | C_1 \cap \dots \cap C_k)$
- *liste des séjours de l'ensemble  $E \cap C_1 \cap \dots \cap C_k$*
- $risque\ relatif\ RR = \frac{P(E / C_1 \cap \dots \cap C_k)}{P(E / (C_1 \cap \dots \cap C_k))}$
- « *p value* » du test exact de Fisher pour d'indépendance entre  $(E)$  et  $(C_1 \cap \dots \cap C_k)$
- *délai médian de survenue de l'effet, une fois toutes les conditions remplies*
- *autres statistiques permettant de caractériser le terrain et les conséquences*

### 2.3.5 Identification et présentation des cas d'EIM

Les cas potentiels d'EIM sont ceux qui d'une part remplissent les conditions d'une règle et d'autre part subissent l'effet décrit par la règle, avec une compatibilité temporelle entre les conditions et l'effet. Afin de restreindre l'affichage aux cas les plus pertinents, un filtrage des règles est réalisé dynamiquement dans chaque service sur les critères suivants : la règle détecte au moins un cas ET le risque relatif est supérieur à 1 ET la confiance de la règle est supérieure à 5% ET la « p value » du test exact de Fisher est inférieure à 5%

Une interface graphique est développée : les « ADE Scorecards ». Lorsqu'un médecin d'un service se connecte à l'outil, il visualise une information contextualisée pour son service. Cela inclut d'une part des informations sur les règles qui détectent des cas d'EIM dans son service, ainsi que les statistiques liées à ces règles et cas, et d'autre part la possibilité de visualiser toute l'information disponible pour chacun des cas, ces cas étant représentés selon le modèle de données défini précédemment.

## 3 Résultats

### 3.1 Exemple d'arbre de décision et d'analyse de règles

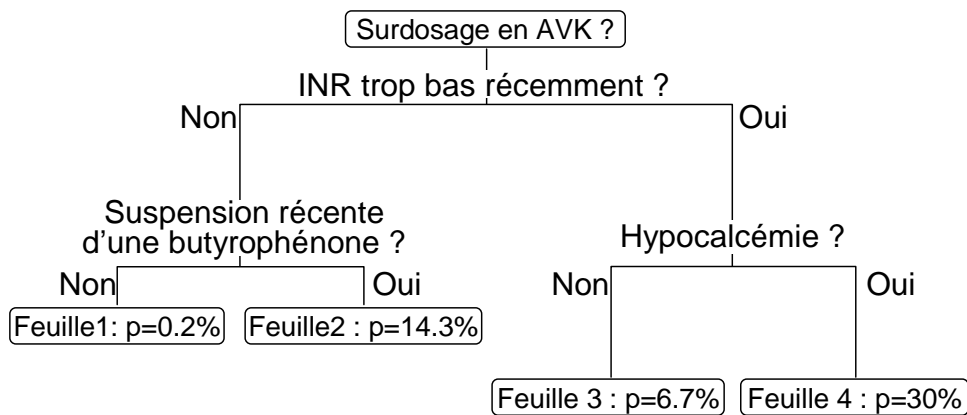


Figure 3 : Exemple d'arbre : quelles sont les circonstances expliquant un surdosage ( $INR \geq 5$ ) en anti-vitamine K (AVK) chez les patients sous AVK ?

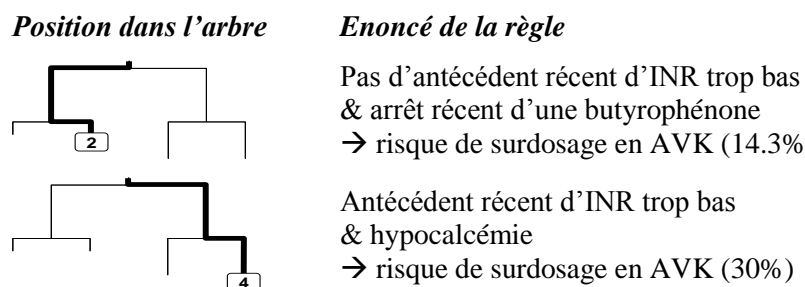


Figure 4 : Exemple d'arbre : détail des 2<sup>o</sup> et 4<sup>o</sup> règles obtenues ci-avant

La figure 3 montre un exemple d'arbre de décision obtenu dans un service particulier. Dans cet exemple, nous cherchons à découvrir les circonstances qui pourraient expliquer la survenue d'un surdosage en anti-vitamine K (AVK). Les AVK sont administrés chez des patients ayant un risque thromboembolique. Les AVK inhibent la coagulation, ce qui a pour effet de réduire ce risque. Un indice biologique, l'INR (International Normalized Ratio), est utilisé pour évaluer la tolérance et l'efficacité du traitement. Lorsque l'INR

descend sous la zone thérapeutique, il existe un risque de thrombose. Lorsque l'INR atteint ou dépasse une valeur de 5, c'est un surdosage en AVK, il existe un risque d'hémorragie.

La figure 4 montre le détail des 2<sup>o</sup> et 4<sup>o</sup> règles obtenues ci-avant. La deuxième règle énonce que, en l'absence d'antécédent d'INR trop bas, la suppression d'un neuroleptique de classe butyrophénone est suivie dans 14.3% des cas d'un surdosage en AVK. Effectivement, ces neuroleptiques stimulent le transit intestinal, ce qui a pour effet de diminuer l'action des AVK. Les doses d'AVK sont probablement augmentées pour compenser et, lors de la suppression du neuroleptique, le transit intestinal ralentit et on assiste à un surdosage en AVK.

Les effets pharmacocinétiques des suppressions de médicaments sont connus des pharmaciens mais presque toujours absents des RCP. La quatrième règle montre que dans ce service, chez les patients qui ont récemment eu un INR trop bas (sous-dosage ou phase de mise en place du traitement) et qui présentent une hypocalcémie, le risque de surdosage en AVK s'élève à 30%. Selon les experts interrogés, l'hypocalcémie est en fait ici un témoin plus pragmatique d'une hypoalbuminémie. Les AVK se lient à l'albumine dans le plasma. Une hypoalbuminémie peut augmenter l'effet des AVK. Or, de manière artificielle, les hypoalbuminémies entraînent une « fausse » hypocalcémie. Le problème est ici que le taux de calcium est fréquemment mesuré, ce qui n'est pas le cas du taux d'albumine. La présente règle est donc très pragmatique, tandis que les RCP, qui se limitent au taux de protéines total moins fréquemment mesuré, présentent plutôt l'état de la connaissance théorique.

### 3.2 Règles de détection d'EIM intégrées dans l'outil

Comme mentionné dans la section *Matériel et Méthodes*, 83 des 228 types d'EIM décrits dans les RCP sont tracés dans les données sous la forme de 56 variables. Les effets non tracés correspondent à des événements purement cliniques comme « nausées » ou « rash », en l'état actuel des données ils ne sont pas présents de manière suffisamment systématisée. Parmi ces 56 variables, 26 permettent de découvrir des règles de détection. Pour 18 autres effets, les méthodes de data mining mettent en évidence des facteurs en relation avec les pathologies du patient plutôt qu'avec l'usage des médicaments. Enfin, 12 types d'effet sont absents ou quasi-absents des données explorées. L'un d'entre eux fait toutefois l'objet de règles ajoutées à la demande expresse des experts.

Les règles incorporées dans la base de connaissance prédisent donc 27 types d'EIM différents. 236 règles sont ainsi décrites dans la base de connaissance. Ces règles peuvent être classées en fonction de l'effet qu'elles prédisent, comme dans le tableau 1. On peut également classer les 236 règles selon leur origine et la connaissance qu'elles apportent.

On définit ainsi 4 catégories :

- **RCP seules:** ces règles sont décrites dans les RCP mais ne sont pas retrouvées par data mining. Dans ce travail, on calcule toutefois des statistiques intéressantes et contextualisées: **40 règles**
- **DM & RCP:** ces règles ont été trouvées par data mining et existent à l'identique ou presque dans les RCP: **25 règles**
- **DM+:** ces règles ont été découvertes par data mining et ajoutent une connaissance supplémentaire par rapport aux RCP: **127 règles**
- **DM++:** ces règles découvertes par le data mining sont absentes des RCP, mais sont expliquées par une connaissance validée (ex. : effet de l'arrêt d'un médicament, hypocalcémie au lieu d'hypoalbuminémie, etc.) **44 règles**

Tableau 1 : Nombre de règles par effet tracé

Effet	# règles
<i>Troubles de l'hémostase</i>	
Hémorragie (détectée par l'administration d'hémostatique)	7
Surdosage en AVK (INR $\geq$ 5 ou prescription de vitamine K)	59
Surdosage en héparine (temps de céphaline activée >1.23)	5
Thrombopénie (plaquettes <75 000)	24
Autres perturbations de la coagulation	23
<i>Troubles ioniques ou rénaux</i>	
Hyperkaliémie (K <sup>+</sup> >5.3 mmol/l)	63
Insuffisance rénale (créatinine >135 $\mu$ mol/l ou urée >16.6 mmol/l)	8
Autres déséquilibres ioniques	4
<i>Divers</i>	
Anémie (Hémoglobine <10g/dl)	2
Infection bactérienne (détectée par l'administration d'antibiotique)	4
Diarrhée (détectée par l'administration d'un anti-diarrhéique)	2
Infection fongique (détectée par l'administration d'un antifongique)	10
Cholestase hépatique (phos. alcaline >240 UI/l ou bilirubine >22 $\mu$ mol/l)	3
Cytolyse hépatique (ASAT ou ALAT >110 UI/l)	4
Hyperéosinophilie (éosinophiles >10 <sup>9</sup> /l)	4
Élévation des enzymes pancréatiques (amylase ou lipase >90 UI/l)	7
Neutropénie (polynucléaires neutrophiles <1,500/mm <sup>3</sup> )	2
<i>Autres effets</i>	5
<i>Total</i>	236

### 3.3 Les ADE Scorecards par l'exemple

Les résultats précédents sont présentés aux médecins des services via un outil de visualisation des connaissances et d'exploration des cas : les « ADE Scorecards » (tableaux de bord sur les EIM). Il s'agit d'un outil web qui accepte en entrée les séjours d'un hôpital décrits selon le modèle de données précédemment explicité, et les règles de détection d'EIM validées par les experts. En retour, cet outil permet de visualiser pour un service donné les règles activées, les statistiques contextualisées et le détail des séjours suspects. Cet outil est ici présenté sous la forme d'un cas d'utilisation :

« Un médecin travaillant dans un hôpital pour lequel l'outil est accessible, se connecte. Premièrement, il souhaite avoir une vue d'ensemble sur les EIM détectés dans son service durant les 6 derniers mois. Deuxièmement, il veut explorer un cas détecté pour forger sa propre opinion sur l'existence de l'EIM. »

Après connexion, l'utilisateur peut choisir la période d'analyse et voir, par effet détecté, mois après mois, le nombre de cas détectés par le système. A partir de cette page de synthèse, il pourra visualiser les tableaux de bord effet par effet. Dans ce cas, l'utilisateur parcourt le tableau de bord des hyperkaliémies (figure 5).

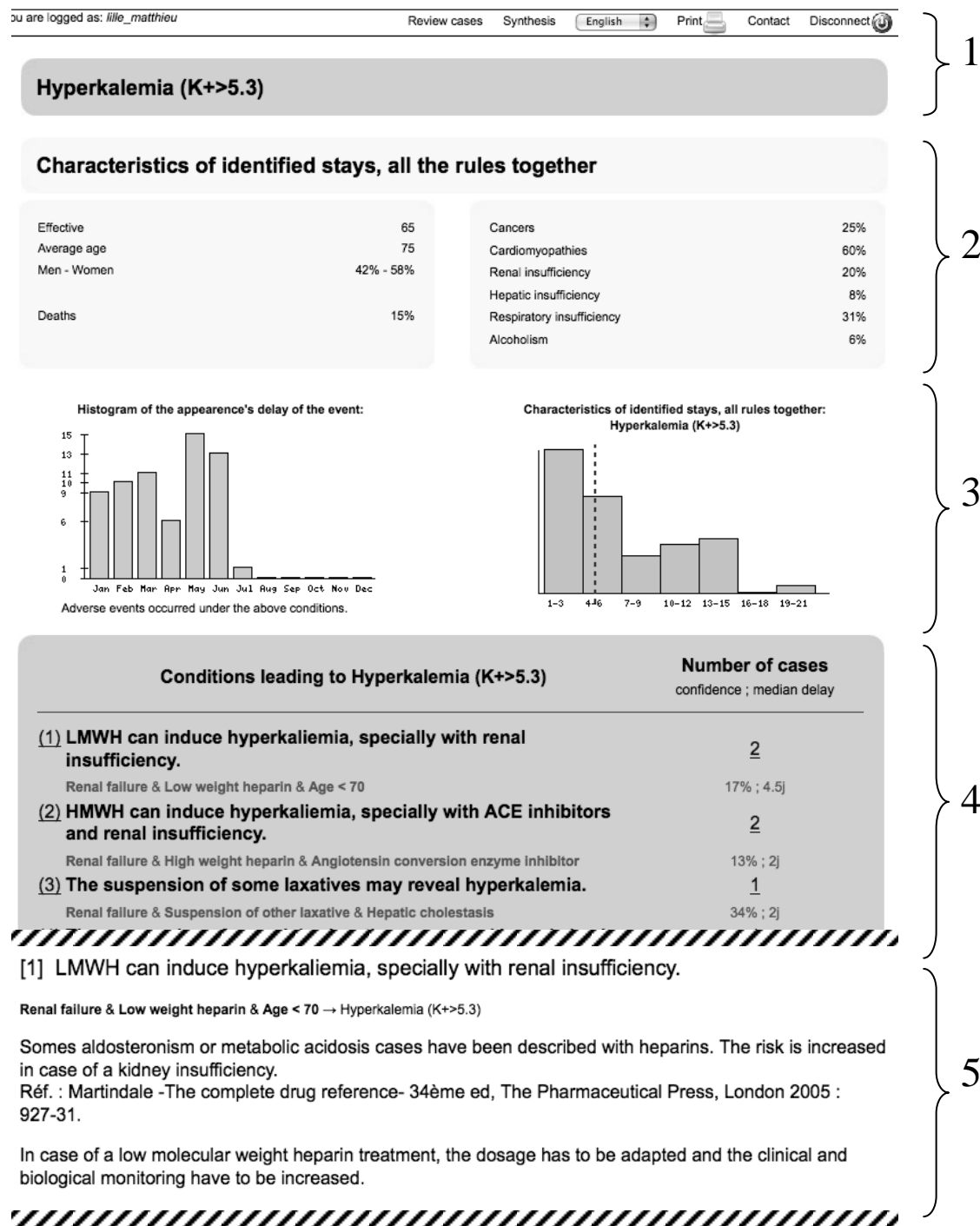


Figure 5 : Tableau de bord des hyperkaliémies (zones hachurées retirées pour réduire l'image)

Ce tableau de bord est constitué de 5 zones (numérotées à droite sur la figure 5) :

1. En haut se trouve le nom du service, la période concernée, et l'effet examiné.
2. Une deuxième zone montre les statistiques disponibles globalement sur tous les cas d'hyperkaliémies détectés, quelle que soit la règle (démographie, pourcentage de décès, proportion de certaines maladies chroniques, etc.).
3. Puis le nombre de cas par mois est représenté, ainsi qu'un histogramme du délai d'apparition de l'effet.



4. Ensuite, toutes les règles utiles sont affichées : intitulé, liste des conditions, nombre de cas détectés, confiance de la règle (proportion d'effet lorsque les conditions de la règle sont remplies).
5. Enfin, pour chaque règle, il est possible de lire les explications détaillées. Des liens hypertexte internes à la page favorisent la lecture, tout en permettant de rester compatible autant avec une utilisation à l'écran qu'avec une impression papier.

Dans la zone numérotée 4, en face de chaque règle se trouve le nombre de cas détectés par la règle. Il suffit de cliquer sur ce nombre pour voir apparaître la liste des séjours suspects (figure 6). Ensuite, un clic sur le numéro de séjour permet de le visualiser en détail. Le séjour s'ouvre immédiatement. Il est possible d'explorer, à travers plusieurs onglets, toute l'information disponible : information administrative, démographique, diagnostics, actes, médicaments, biologie et courriers en texte libre.

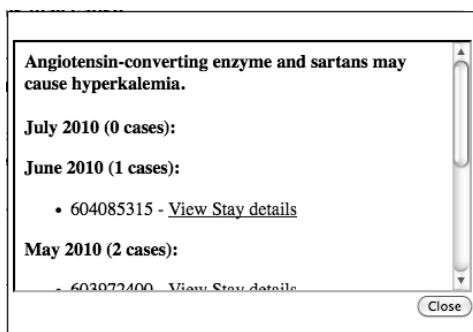


Figure 6 : Accès à la liste des séjours

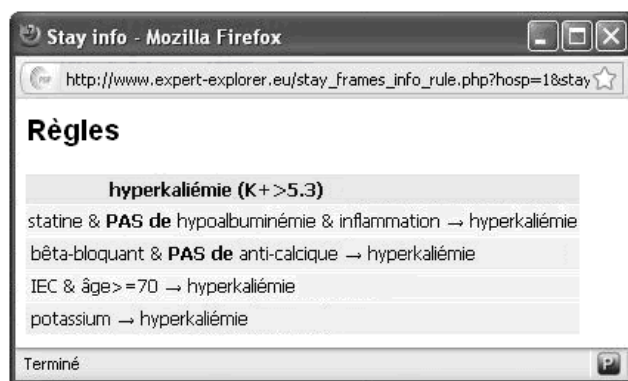


Figure 7 : Visualisation des règles qui se déclenchent sur le séjour

ATC	Nom du médicament	0	5	10
	1000 ml POLYIONIQUE G5 + POTASSIUM CHL.A + SPASFON SOL INJ	■		
A02BC02	INPOMP 20 MG, CPR (VOIR INEXIUM 20)	■■■■■		
A06AD15	FORLAX 100 PDR ORALE SACHET PR SOL BUV	■■■■■		
A07DA03	IMODIUM 2MG GELULE VERT FONCE ET VERT CLAIR	■		
A07XA04	TIORFAN 100MG GELULE BLEU CLAIR ET BLEU FONC	■■		
A10AB05	NOVOMIX 30 FLEXPEN 100 UIML, SUSP INJ, STYLO 3 M	■■■■■		
A12BA01	KALEORID 1000MG LP CPR ENR	■■■■■		
B01AC06	KARDEGIC 75 M0, PDR PR SOL BUV, SACHET	■■■■■		
C01DX12	CORVASAL 4MG CPR SECABLE BLANC	■■■■■		
C07AB03	ATENOLOL 50 MG ARROW, CPR	■■■■■		
C09BA04	PRETERAX, CPR	■■■■■		
C10AA03	ELISOR 20MG CPR SECABLE = PRAVASTATINE 20 MG	■■■■■		
J01DA23	OROKEN 200MG CPR PELLICULE (PRESCRIPTION PAPER, OBLIGATOIRE ANTIBIO)	■■■■■		
N06DX02	TANAKAN 40MG CPR ENR BRUN ROUGE	■■■■■		
V03AE01	KAYEXALATE 4540 PDR PR SUSP BUVIRECT	■		

Figure 8 : Exploration d'un séjour en détail

Nous montrerons simplement ici la page synthétique représentant les médicaments et la biologie (figure 8). Sur le panneau de droite (étiquette 1), pour afficher un graphique de biologie il suffit de cocher la case correspondante. Les éléments intéressants sont d'emblée affichés sur fond de couleur. Ici nous avons affiché le potassium (étiquette 2) qui atteint la valeur de 5,7 mmol/l au 7<sup>ème</sup> jour. En cliquant sur un des boutons du haut de l'écran, une popup rappelle les règles impliquées (figure 7). Dans le volet gauche de l'application (figure 8) on peut contrôler que les médicaments incriminés sont bien administrés avant la survenue de l'hyperkaliémie, ils apparaissent automatiquement sur fond de couleur. On peut vérifier la présence de potassium (étiquette 3), de bêtabloquant (étiquette 4), de l'association d'IEC et de diurétique thiazidique (étiquette 5), et de statine (étiquette 6) avant le 7<sup>ème</sup> jour. Tous ces médicaments sont hyperkaliémifiants. Sur cet exemple, l'utilisateur peut même voir la réaction des prescripteurs. Heureusement, le potassium n'avait été administré que temporairement (étiquette 3) mais, comme la kaliémie est très élevée au 7<sup>ème</sup> jour, un chélateur du potassium est immédiatement administré (étiquette 5). Dans ce cas précis, l'utilisateur pourra vérifier que l'hyperkaliémie n'a été mentionnée ni dans les diagnostics CIM10, ni dans le courrier de sortie.

#### 4 Discussion et conclusion

Les *ADE Scorecards* sont un outil innovant en matière de détection et de présentation des EIM. Pour ce qui concerne les règles de détection des EIM, le formalisme retenu ici est classique, de même que le fait d'utiliser les médicaments administrés, les variables démographiques et les résultats de biologie [2, 15-22]. En revanche il est unique d'associer indifféremment tous ces types de variables, chaque système étant le plus souvent limité à 1 ou 2 types simultanément. De plus, ce travail introduit des types de conditions jusque là ignorées : la suspension d'un médicament, et des facteurs liés au déroulement du séjour (exemple : antécédent d'INR trop bas). La plupart des règles produites ici rentrent dans la catégorie « AA5 – Advanced alerts – Complex prescribing alerts » définie par [22]. Enfin, si des données d'autres types étaient systématiquement disponibles, comme par exemple des résultats d'ECG, la méthode employée ici n'aurait aucune difficulté à les intégrer. Pour ce qui concerne la méthode de production des règles, présentement le data mining, il n'existe pas dans la littérature d'exemple similaire. Le data mining est fréquemment employé sur les rapports d'EIM, mais jamais semble-t-il sur les dossiers électroniques eux-mêmes. Par cette approche nous évitons le biais lié à la forte sous-déclaration des EIM. La méthode de génération des règles souffre cependant de deux faiblesses notables. Tout d'abord, le data mining ne peut pas identifier des situations qui ne sont pas observables dans les données, c'est la raison pour laquelle nous avons dans ce travail intégré manuellement des règles issues des RCP, principalement des contre-indications absolues. Ensuite, il n'est pas possible de détecter et d'exploiter des événements qui ne sont pas explicitement décrits dans les données, comme certains événements cliniques (nausées, rash, etc.). Le présent travail est poursuivi et une mesure prospective d'impact est menée au centre hospitalier de Denain (France). Le but est d'évaluer l'appréciation de l'outil par les médecins, et de mesurer si son déploiement a un impact sur les pratiques de soins, voire sur le nombre d'EIM.

#### Remerciements

La recherche aboutissant à ces résultats a été financée par le 7<sup>o</sup> programme cadre de la Communauté Européenne (FP7/2007-2013), sous l'agrément n°216130-*the PSIP project*.

## Références

- [1] Kohn LT, Corrigan JM, Donaldson MS. *To Err Is Human: Building a Safer Health System*. Washington, DC: National Academy Press; 1999: 287 p.
- [2] Morimoto T, Gandhi TK, Seger AC, Hsieh TC, Bates DW. Adverse drug events and medication errors: detection and classification methods. *Qual Saf Health Care*. 2004; 13(4): 306-14.
- [3] Murff HJ, Patel VL, Hripcsak G, Bates DW. Detecting adverse events for patient safety research: a review of current methodologies. *J Biomed Inform* 2003; 36(1-2): 131-43.
- [4] Aramaki E, Miura Y, Tonoike M, Ohkuma T, Masuichi H, Waki K, Ohe K. Extraction of adverse drug effects from clinical records. *Stud Health Technol Inform*. 2010; 160: 739-43.
- [5] Almenoff J, Tønning JM, Gould AL, Szarfman A, Hauben M, Ouellet-Hellstrom R et al. Perspectives on the use of data mining in pharmaco-vigilance. *Drug Saf*. 2005; 28: 981-1007.
- [6] Bate A, Edwards IR. Data mining in spontaneous reports. *Basic Clin Pharmacol Toxicol*. 2006; 98(3): 324-30.
- [7] Chazard E, Ficheur G, Merlin B, Serrot E, PSIP Consortium, Beuscart R. Adverse drug events prevention rules: multi-site evaluation of rules from various sources. *Stud Health Technol Inform*. 2009; 148: 102-11.
- [8] Pereira S, Sakji S, Névéal A, Kergourlay I, Kerdelhué G, Serrot E, Joubert M, Darmoni SJ. Multi-terminology indexing for the assignment of MeSH descriptors to medical abstracts in French. *AMIA Annu Symp Proc*. 2009;:521-5.
- [9] Date CJ. *Database in Depth: Relational Theory for Practitioners*. O'Reilly (2005).
- [10] Chazard E, Merlin B, Ficheur G, Sarfati JC, PSIP Consortium, Beuscart R. Detection of adverse drug events: proposal of a data model. *Stud Health Technol Inform*. 2009; 148: 63-74.
- [11] Breiman L. *Classification and regression trees*. Belmont, Calif.: Wadsworth International Group; 1984.
- [12] Zhang HP, Crowley J, Sox H, Olshen RA. *Tree structural statistical methods*. In : *Encyclopedia of Biostatistics*. Chichester, England: Wiley; 2001. p. 4561-73.
- [13] Therneau TM, Atkinson B, Ripley BD. Rpart: Recursive Partitioning. Disponible sur <<http://CRAN.R-project.org/package=rpart>> (Consulté le 26.12.2010)
- [14] Agrawal R, Imielinski T, Swami A, editors. Mining Association Rules between Sets of Items in Large Databases. Proceedings of the ACM SIGMOD International Conference on Management of Data; 1993 May. Washington D.C.
- [15] Bates DW, O'Neil AC, Boyle D, Teich J, Chertow GM, Komaroff AL, Brennan TA. Potential identifiability and preventability of adverse events using information systems. *JAMIA* 1994; 1(5): 404-11.
- [16] Jha AK, Kuperman GJ, Teich JM, Leape L, Shea B, Rittenberg E, Burdick E, Seger DL, Vander Vliet M, Bates DW. Identifying adverse drug events: development of a computer-based monitor and comparison with chart review and stimulated voluntary report. *JAMIA* 1998; 5(3): 305-14.

- [17] Kuperman GJ, Teich JM, Bates DW, Hiltz FL, Hurley JM, Lee RY, Paterno MD. Detecting alerts, notifying the physician, and offering action items: a comprehensive alerting system. *Proc AMIA Annu Fall Symp* 1996;: 704-8.
- [18] Del Fiol G, Rocha BH, Kuperman GJ, Bates DW, Nohama P. Comparison of two knowledge bases on the detection of drug-drug interactions. *Proc AMIA Symp*. 2000;: 171-5.
- [19] Gandhi TK, Weingart SN, Seger AC, Borus J, Burdick E, Poon EG, Leape LL, Bates DW. Outpatient prescribing errors and the impact of computerized prescribing. *J Gen Intern Med* 2005; 20(9): 837-41.
- [20] Judge J, Field TS, DeFlorio M, Laprino J, Auger J, Rochon P, Bates DW, Gurwitz JH. Prescribers' responses to alerts during medication ordering in the long term care setting. *J Am Med Inform Assoc* 2006; 13(4): 385-90.
- [21] Paterno MD, Maviglia SM, Gorman PN, Seger DL, Yoshida E, Seger AC, Bates DW, Gandhi TK. Tiering drug-drug interaction alerts by severity increases compliance rates. *JAMIA* 2009; 16(1): p. 40-6.
- [22] Schedlbauer A, Prasad V, Mulvaney C, Phansalkar S, Stanton W, Bates DW, Avery AJ. What evidence supports the use of computerized alerts and prompts to improve clinicians' prescribing behavior? *JAMIA* 2009; 16(4): 531-8.
- [23] Teich JM, Glaser JP, Beckley RF, Aranow M, Bates DW, Kuperman GJ, Ward ME, Spurr CD. The Brigham integrated computing system (BICS): advanced clinical systems in an academic hospital environment. *Int J Med Inform* 1999; 54(3): 197-208.
- [24] Honigman B, Lee J, Rothschild J, Light P, Pulling RM, Yu T, Bates DW. Using computerized data to identify adverse drug events in outpatients. *JAMIA* 2001; 8(3): 254-66.
- [25] Field TS, Gurwitz JH, Harrold LR, Rothschild JM, Debellis K, Seger AC, Fish LS, Garber L, Kelleher M, Bates DW. Strategies for detecting adverse drug events among older persons in the ambulatory setting. *J Am Med Inform Assoc* 2004; 11(6): 492-8.

### **Adresse de correspondance**

Emmanuel Chazard, Service d'Information et des Archives Médicales, Clinique de Santé Publique, CHRU de Lille, 150 rue du Dr Yersin, 59037 Lille Cedex, France.

Courriel : emmanuel@chazard.org