




ELSEVIER
MASSON

Disponible en ligne sur
 ScienceDirect
www.sciencedirect.com

Elsevier Masson France
EM|consulte
www.em-consulte.com

IRBM

IRBM xxx (2009) xxx-xxx

Communication brève

Détection et prévention des effets indésirables liés aux médicaments par *data-mining*

Data-mining-based detection and prevention of adverse drug events

E. Chazard^{a,*}, C. Preda^c, B. Merlin^b, G. Ficheur^b, R. Beuscart^{a,b}

^a Centre d'études et de recherche en informatique médicale, EA2694, université Lille II, Lille, France

^b Secteur d'informatique et information médicale, CHRU de Lille, Lille, France

^c Laboratoire de mathématiques Paul-Painlevé, université Lille I, Lille, France

Reçu le 16 avril 2009 ; accepté le 26 mai 2009

Résumé

Les effets indésirables liés aux médicaments causeraient 10 000 décès par an en France. Le plus souvent, la détection de ces effets repose sur les déclarations et l'écriture de règles d'alerte et de prévention est faite à dire d'expert lors de revues des dossiers. De plus, les spécificités des services ne sont pas prises en compte (patients, pratiques, connaissances). L'objectif du projet européen PSIP est d'utiliser la *data-mining* pour détecter ces effets et produire automatiquement les règles de contrôle, service par service. Sur 10 500 séjours danois et français, nous obtenons à ce jour 630 règles dont 75 sont validées. L'article présente un exemple d'arbre de décision et l'interprétation autour des précautions d'emploi des antivitamine K. L'exploitation des résultats ne s'entend qu'en contexte. Un procédé similaire pourrait être utilisé dans d'autres domaines.
© 2009 Elsevier Masson SAS. Tous droits réservés.

Mots clés : Arbres de décision ; *Data-mining* ; Effets indésirables des médicaments ; Informatique médicale

Abstract

Adverse drug events would be responsible from 10,000 death per year in France. Most often, detection relies on events declaration; alert and prevention rules are written by experts thanks to cases reviews. Moreover, the medical departments' specificities are not taken into account (patients, practices, knowledge). The objectives of the PSIP European project is to use data-mining to detect those events and automatically generate control rules, department by department. Using 10,500 French and Danish stays, we obtain till now 630 rules from which 75 have been validated. The article shows an example of a decision tree and its interpretation in the field of cautions taken during vitamin K antagonists' use. Results' exploitation cannot elude contextualization. A similar process may have many different uses.
© 2009 Elsevier Masson SAS. All rights reserved.

Keywords: Adverse drug events; Data-mining; Decision trees; Medical informatics

1. Introduction

Chaque année, les effets indésirables médicamenteux seraient responsables de 10 000 décès en France et 98 000 aux États-Unis [1]. En milieu hospitalier, une prévention est possible lors de la prescription du médicament lorsque

cette prescription est connectée et s'intègre dans un circuit du médicament informatisé. On peut alors implémenter des règles d'alerte, par exemple en cas de prescription malgré une contre-indication. Toutefois, ces règles sont écrites par des experts, appuyés sur la connaissance académique elle-même fondée sur des déclarations insuffisantes [2,3]. De plus, les règles ainsi produites s'appliquent à l'identique dans des services pourtant différents : les patients et la connaissance des produits sont différents. Les alertes sont trop nombreuses et peu pertinentes. L'objectif est ici de produire automatiquement la connaissance par *data-mining* [4], service par service.

* Auteur correspondant.

Adresse e-mail : emmanuelchazard@yahoo.fr (E. Chazard).

2. Matériel et méthodes

Une première phase a consisté à proposer un modèle de données commun aux différents partenaires fournisseurs de données (hôpitaux régionaux de Copenhague, CHRU de Lille, CHU de Rouen, centre hospitalier de Denain) et à automatiser l'extraction des données. Point d'importance, il ne s'agissait pas d'organiser un recueil spécifique, mais d'exploiter des données déjà recueillies en routine [5,6] : informations sur le parcours (programme de médicalisation des systèmes d'information, PMSI), diagnostics médicaux, mesures de biologie, prescriptions médicamenteuses.

La seconde phase du projet est l'exploitation statistique de ces données. Nous avons d'abord développé des moteurs d'agrégation alimentés par des *mappings* (Fig. 1). Une agrégation des données est nécessaire, afin de transformer des encodages complexes en notions binaires. Quelques exemples sont :

- l'agrégation des diagnostics : « insuffisance rénale vaut 1 si le séjour contient au moins un des codes suivants... » ;
- l'agrégation de la biologie : « apparition d'une cytolysé hépatique vaut 1 si le motif suivant se produit... » ;
- l'agrégation des prescriptions : « prescription d'un inhibiteur enzymatique vaut 1 si au moins un des médicaments suivants est prescrit... ».

Cette agrégation permet à ce jour de générer automatiquement 55 variables de possible effet indésirable médicamenteux (variables à expliquer) et 576 variables de cause ou contexte de cet effet (variables explicatives). Nous utilisons ensuite des

arbres de décision [7-12] afin d'établir le cas échéant un lien statistique entre les causes ou contexte et les effets indésirables potentiels. Les analyses sont faites à l'aide du logiciel R, en utilisant l'algorithme Cart. Les arbres sont interprétés par des experts et comparés à la connaissance bibliographique (résumé des caractéristiques des produits, articles scientifiques, sociétés savantes de pharmacologie).

3. Résultats

Tandis que l'extraction de données se poursuit, des résultats provisoires ont été obtenus sur 10 500 séjours issus de quatre services français et danois. La fouille de données génère à ce jour près de 630 règles, dont 75 sont déjà validées. Les règles sont très différentes en fonction des services : les pratiques, les patients et la maîtrise du risque varient.

Dans cet exemple, nous traçons l'effet « survenue d'un *international normalized ratio* (INR) trop bas ». Un INR trop bas expose le patient sous antivitamine K (AVK) à un risque de thrombose. Un arbre est automatiquement généré à partir des 280 variables potentiellement explicatives. La première bifurcation indique que la variable la plus liée à la survenue d'un INR trop bas est l'existence d'un INR trop élevé à l'admission (risque hémorragique). Lorsqu'un patient entre dans ce service avec un INR trop élevé (hémorragie probable), dans 29 % des cas le praticien surcorrige la prescription d'AVK, et l'INR devient trop faible (Fig. 2). Les patients âgés admis pour INR trop élevé et présentant une hypoalbuminémie sont surcorrigeés dans 87 % des cas (Fig. 3). L'hypoalbuminémie est probablement la cause de l'INR trop élevée et n'est pas suffisamment prise en compte lors de la correction de dose. En poursuivant sur les branches

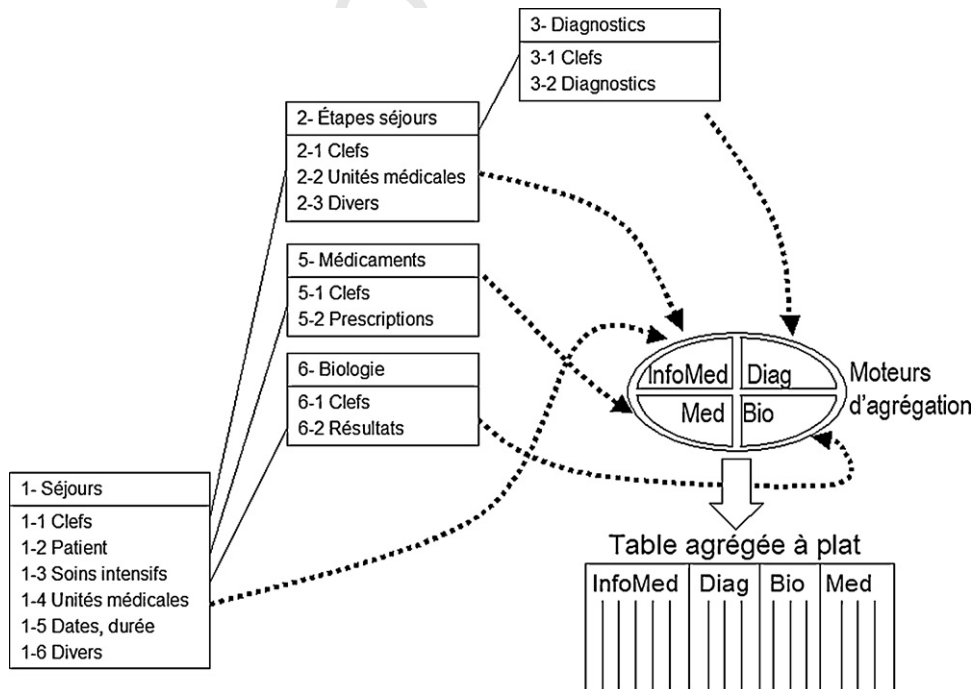


Fig. 1. Le parcours, les diagnostics, les prescriptions médicamenteuses et les mesures de biologies sont agrégées en 576 variables explicatives et 55 variables à expliquer.

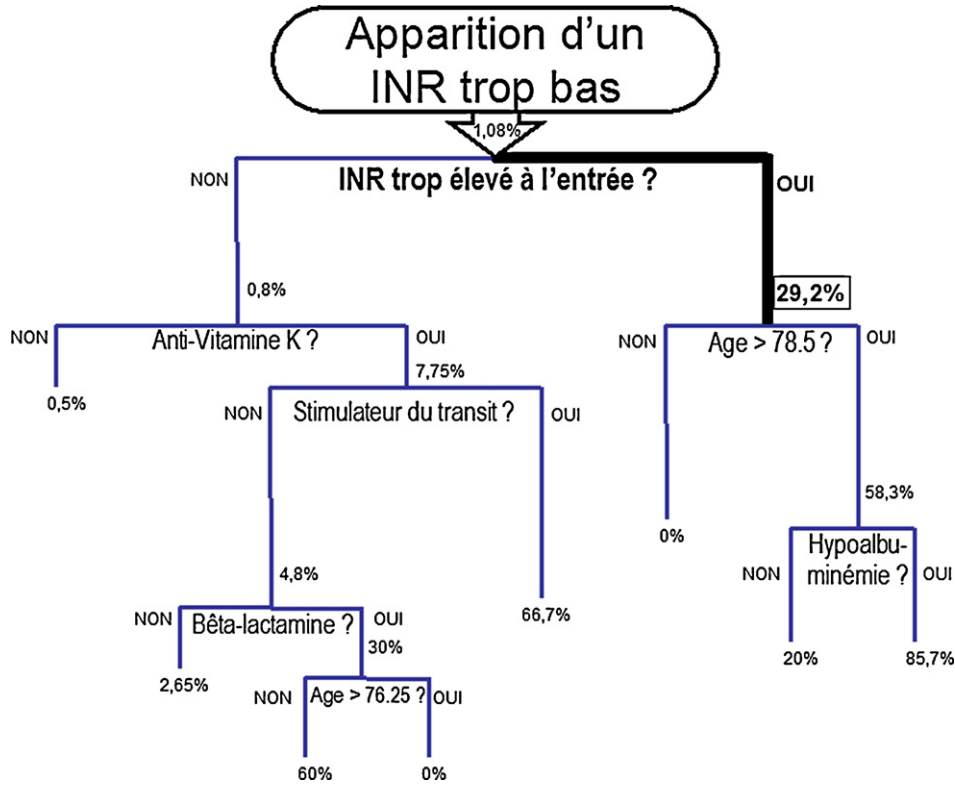


Fig. 2. Lorsque le patient est admis avec un INR trop élevé, un INR trop bas survient dans 9 % des cas.

100 de gauche, les patients entrant avec un INR normal et recevant
101 un AVK et un stimulateur du transit intestinal subissent un INR
102 trop bas dans 67 % des cas (Fig. 4). Les stimulateurs du transit
103 diminuent la biodisponibilité des AVK. De même, les patients de

moins de 76 ans auxquels sont prescrits un AVK et une bêta-lactamine subissent dans 60 % des cas un INR trop bas (Fig. 5). Les infections augmentent le catabolisme hépatique, ce qui diminue la biodisponibilité des AVK.

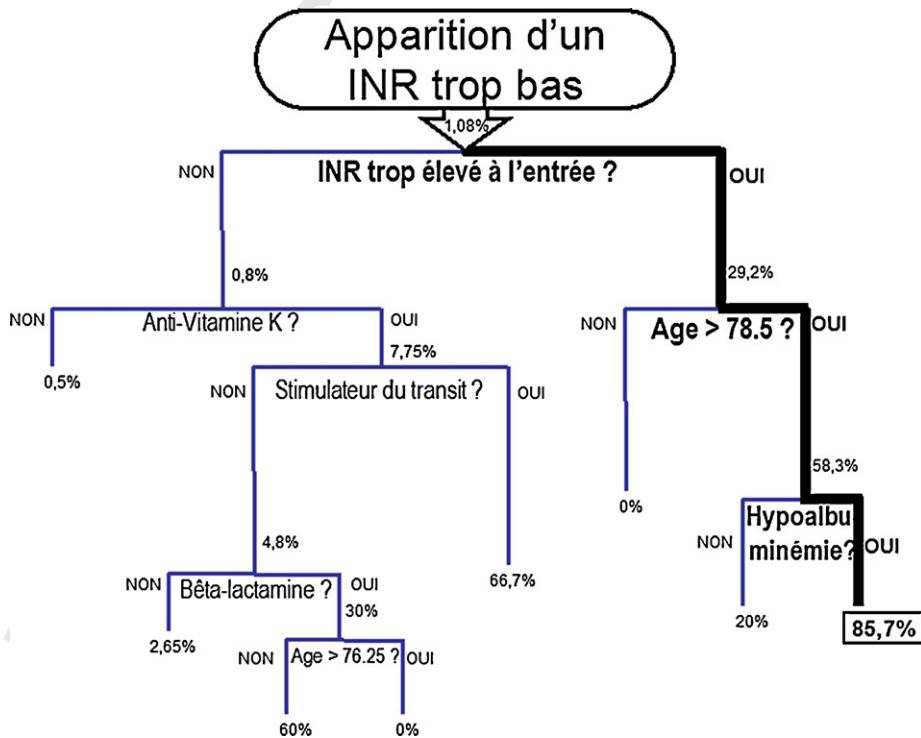


Fig. 3. Patient de plus de 78 ans admis avec un INR trop élevé et une hypoalbuminémie : survenue d'un INR trop bas dans 86 % des cas.

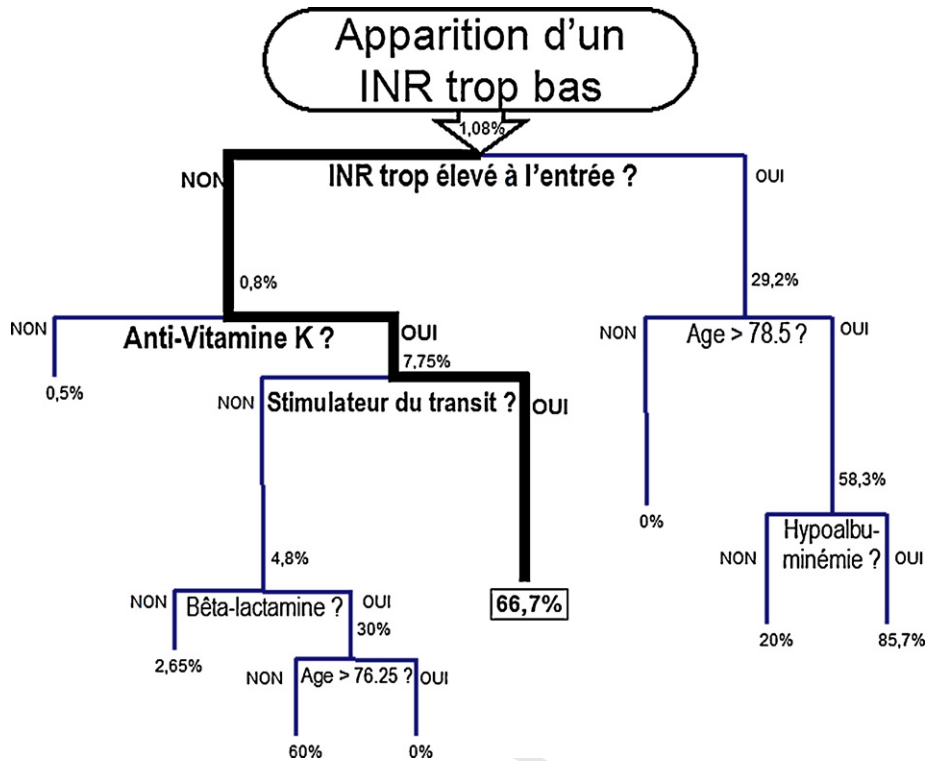


Fig. 4. Antivitamine K et stimulateur du transit : survenue d'un INR trop bas dans 67 % des cas.

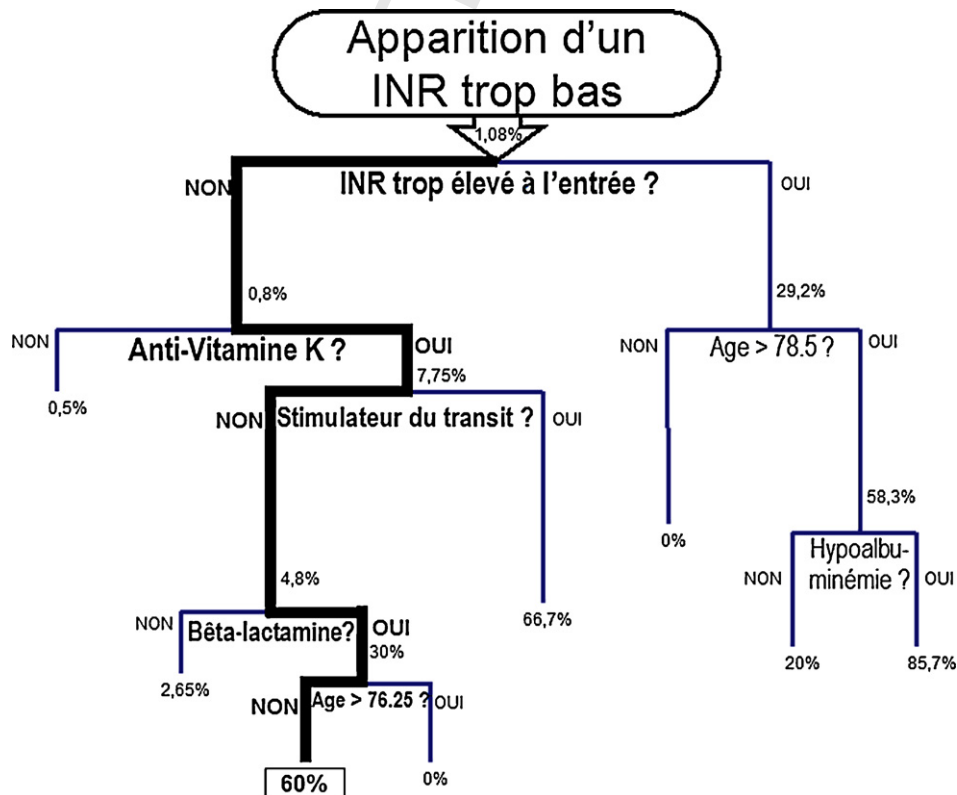


Fig. 5. Antivitamine K et bêtalactamine et âge inférieur à 76 ans : survenue d'un INR trop bas dans 60 % des cas.

Les précautions d'emploi décrites ci-dessus sont connues et enseignées. Néanmoins, elles sont noyées dans une longue liste, alors que seule une petite partie de cette liste serait utile au présent service. Les arbres des autres services sont très différents.

4. Discussion et conclusion

Les premiers résultats sont encourageants [12] et préfigurent une nouvelle approche de l'effet indésirable médicamenteux, les approches actuelles étant essentiellement tournées vers les revues de cas [13] ou la fouille de données sans utilisation de méthode statistique [14-16]. De plus, les premiers résultats montrent à quel point les mêmes règles ne peuvent s'appliquer dans des services différents : la contextualisation est indispensable.

En amont de la fouille de données, les améliorations du processus portent sur les points suivants :

- l'amélioration des moteurs d'agrégation des diagnostics, des médicaments et des paramètres biologiques ;
- la prise en compte de la cascade d'événements : un premier effet indésirable (insuffisance rénale) peut participer à un second (par surdosage d'un autre médicament), cette cascade doit être prise en compte.

Pour ce qui est de la fouille stricto sensu, des méthodes concurrentes aux arbres de décision doivent être éprouvées.

Les règles sont validées à la lecture des sources académiques. Il est prévu de tester le recoupement entre les règles issues du *data-mining* et les règles proposées par la société Vidal.

En aval de la fouille de données, sur les règles validées, des dossiers sélectionnés par les règles doivent être validés par des experts pharmacologues, c'est l'objet des étapes suivantes. L'utilisation des règles devra ensuite prendre en compte les facteurs humains : au moment d'émettre une alerte, il faut savoir quels leviers a l'utilisateur pour intervenir, en fonction de son profil professionnel et du stade auquel l'alerte est produite.

En perspectives, le modèle de données défini dans *Patient Safety Through Intelligent Procedures in Medication* (PSIP), qui utilise des données souvent recueillies en routine, pourrait être utilisable dans d'autres projets de fouilles de données. L'approche générale utilisée ici pourrait s'avérer utile dans des domaines très différents.

Remerciements

La recherche menant à ces résultats a été financée par le septième programme cadre de la commission européenne (FP7/2007-2013) sous la référence n° 216130. Ce travail s'inscrit dans le projet PSIP.

Références

- [1] Kohn LT, Corrigan JM, Donaldson MS. To err is human. Washington DC: National Academy Press; 1999.
- [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;13:306-14.
- [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;36:131-43.
- [4] Adriaans P, Zantige. *Data Mining*. Edingburgh: Addison Wesley; 1996.
- [5] Gurwitz JH, Field TS, Harrold LR, et al. Incidence and preventability of adverse drug events among older persons in the ambulatory setting. *JAMA* 2003;289:1107-16.
- [6] Jalloh OB, Waitman LR. Improving computerized provider order entry (CPOE). Usability by data mining Users' queries from Access Logs. 2006. p. 379-83.
- [7] Zhang HP, Crowley J, Sox H, Olshen RA. Tree structural statistical methods. *Encyclopedia of Biostatistics*, vol. 6. Chichester, England: Wiley; 2001. p. 4561-73.
- [8] Breiman L, Friedman JH, Olshen R, Stoney C. *Classification and regression trees*. Belmont, California: Wadsworth; 1984.
- [9] Quinlan JR. Introduction of decision trees. *Mach Learn* 1986;1:81-106.
- [10] Fayyad UM, Piatetsky-Shapiro, Smyth P. From data mining to knowledge discovery: an overview. In: *Advances in knowledge discovery and data mining*. MIT Press; 1996. p. 1-36.
- [11] Lavrac N. Selected techniques for data mining in medicine. *Artif Intell Med* 1999;16:3-231.
- [12] Beuscart R, Beuscart-Zéphir MC, PSIP Consortium. Workshop on patient safety through intelligent procedures in medication. 2008.
- [13] Bates DW, Evans RS, Murff H, Stetson PD, Pizziferi L, Hripcsak G. Detecting adverse events using information technology. *J Am Med Inform Assoc* 2003;10:115-28.
- [14] Seger AC, Jha AK, Bates DW. Adverse drug event detection in a community hospital utilising computerised medication and laboratory data. *Drug Saf* 2007;30:817-24.
- [15] Honigman B, Lee J, Rotschild J, Light P, Pulling RM, Yu T, Bates DW. Using computerized data to identify adverse drug events in outpatients. *J Am Med Inform Assoc* 2001;8:254-66.
- [16] Honigman B, Light P, Pulling RM, Bates DW. A computerized method for identifying incidents associated with adverse drug events in outpatients. *Int J Med Inform* 2001;61:21-32.