Adverse Drug Events Detection by Data Mining of Electronic Health Records

Emmanuel Chazard[a], Grégoire Ficheur[a], Sanne Jensen[b], Peter Mc Nair[b,c], Régis Beuscart[a]

[a] *Medical Information and Records Department EA2694, University Hospital, Lille, France*
[b] *Health Informatics Unit, Corporate IT, Capital Region of Denmark*
[c] *Medical Genetics Laboratory Center, Kennedy Center, Glostrup, Denmark*

Abstract and Objective

Adverse drug events (ADEs) are a public health issue. Their detection usually relies on spontaneous declarations and staff-operated reviews. The present work aims at (1) mining electronic health records to automatically identify ADEs (2) generating alert rules to prevent ADEs. Those rules will then be implemented into a clinical decision support system. Methodology: (1) data are aggregated in order to simplify their structure and to take the time into account. (2) cause-to-effect relationships are identified thanks to statistical methods. Results: data-mining of 55,000 hospital stays allows getting 255 validated rules in the field of anti-thrombotic agents and related ADEs. Other ADE-detection rules can be imported into the same repository and automatically evaluated in different medical departments. The results show the importance of segmentation and contextualization of the rules.

Keywords:

Adverse drug events, Patient safety, Data mining, Decision trees, Association rules, Electronic health records.

Introduction

Each year, adverse drug events (ADEs) are responsible from 98,000 deaths in the USA. Clinical decision support system (CDSS) can be used to prevent ADEs thanks to alert rules. Those rules are usually generated by experts. They encounter several problems: lack of coverage due to ADE under-declaration, lack of contextualization needed by the heterogeneity of practices and risk monitoring, lack of segmentation although hazards generally involve several cofactors.

Our objective is to data-mine electronic health records (EHRs) in order to discover ADEs and to generate decision rules for ADE prevention. Those rules are automatically weighted in each medical department. The ADE detection must be automated and fed by routine EHR datasets. In those datasets, the ADEs are not explicitly flagged and no staff-operated review is performed. For that reason, classical machine learning approaches cannot be used.

Material and methods

Data are extracted from the partners' EHRs, including demographic information, diagnoses (ICD10 codes), drug prescriptions (ATC codes) and lab results (C-NPU codes) according to a common data model. 55,000 different hospital stays are mined: 24,000 stays from the Capital Region of Denmark hospitals, and 29,000 stays from Denain, Rouen and Lille hospitals in France. The data are aggregated into sets of events.

Then decision trees and association rules are used in order to automatically discover some associations between unexpected events (such as lab abnormalities) and set of causes, including drugs and drug suppressions but also demographic variables, lab-related context and organizational causes. The associations are filtered and validated by pharmacologists. When those associations fit the academic knowledge, they can be considered as ADE-prevention rules.

Finally, all the rules are stored in a rule repository. This helps automatically computing the confidence of all the rules in every medical department. At this step, other rules can be added and automatically tested.

Results

255 decision rules are discovered and validated, mainly in the field of anticoagulation: conditions leading to INR deviations, hyperkalemia, APPT elevation, thrombopenia, etc. For each rule, in each of the 15 available medical departments, the following statistics are automatically computed: confidence (positive predictive value), support, relative risk, median delay of appearance of the event, Fisher's exact test p value. Pharmacokinetics and demographic factors matter a lot.

Conclusion

Data mining of EHR is a new and useful approach in the field of ADEs. The rules will be implemented into a contextualized CDSS, as the confidences of the rules vary a lot over the different medical departments (monitoring policies).