R-49

Automated monitoring of antithrombotic-related adverse events in electronic medical records using structured data mining and natural language processing

C.Coumau[1,2,3], F. Gaspar[1,2,3] , Z. Schillaci[4], MA Le Pogam[5,] C.Csajka[1,2,3] and the SwissMADE collaborators

[1] Center for Research and Innovation in Clinical Pharmaceutical Sciences, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland.
[2] Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, University of Lausanne, Geneva, Switzerland.
[3] School of Pharmaceutical Sciences, University of Geneva, Geneva, Switzerland.
[4] Effixis SA, AI consulting firm, St-Sulpice, Switzerland
[5] Unisanté Center for Primary Care and Public Health, Department of Epidemiology and Health Systems, University of Lausanne, Lausanne. Switzerland

Purpose

- **About 30% of older patients** experience at least **one adverse drug event** (ADE) during their hospital stay.
- **Antithrombotics are one of the most prescribed and at-risk drugs** for ADEs in older patients.
- **The detection of adverse drug events could be improved by using recent computer technology like named entity recognition (NLP) with the use of the named entity recognition and clustering.**

Conclusion

- Despite many challenges in data collection and harmonization, we have developed a multi-source and multi-modal pipeline to identify antithrombotic-related hemorrhage in a large set of patient stays using an innovative combination of rule-based and natural language processing methodologies.
- The first validation shows that each data source provides complementary information, which needs to be further validated through a manual review of computerized patient records.

Aim

To identify antithrombotic-related hemorrhages detection based on data collected from patients' electronic medical records (EMR) to improve the monitoring of such events in hospital settings.

Methods

Multicenter cross-sectional study using EMR data from the **CHUV** (Lausanne), **HUG** (Geneva) **KSB** (Baden) and **USZ** (Zürich).

Information sources:
- ✓ Structured data: ICD-10[1] /CHOP[2] codes; Laboratory data; Prescription orders
- ✓ Free-text data: Discharge letters

- **Identification** of all variables of interest related to hemorrhages identification.
- **Data extraction**, cleaning and harmonization between the four hospitals
- **Development of detection algorithms of hemorrhages** using **structured data** and **free-text data** (only CHUV) using rule-based algorithms for structured data and natural language processing (NLP) for free texts.
- **Iterative analyses** of (i) the separate contribution of each source of information and (ii) a combination of relevant sources of information, with a final set of 7 algorithms to be tested. The union of the hemorrhages detected by at least one algorithm and the intersection of hemorrhages detected by all algorithms was finally performed to estimate their relative contribution to the overall ADE identification.
- **Validation of the algorithms** using a corpus of patient stays (CHUV) with a positive event of hemorrhages identified by a semi-automatic NLP annotation method and comparison with hemorrhages detected from structured data algorithms to estimate the recall of the different source of information embedded in the algorithms.

[1] International Statistical Classification of Diseases and Related Health Problems 10th Revision – German modification; [2] Swiss Classification of Surgical Procedures

Results

Table 1. Summary of harmonized data in the SwissMADE common database (**37'079 hospital stays**)

Source of information	number of data
Laboratory data	7'874'831
Prescription orders (drugs and procedure)	11'319'477
Clinical measurements	5'579'426
Diagnosis and procedures (ICD-10/CHOP codes)	740'883

Table 2. Hemorrhages detected by structured data algorithms

Information source	number of events (n, %)	
ICD-10 / CHOP codes	4'875	13 %
Laboratory data and prescription orders	9'962	27 %
ICD-10 / CHOP codes U Laboratory data and prescription orders	12'421	33 %
ICD-10 / CHOP codes ∩ Laboratory data and prescription orders	2'416	7 %

U: algorithms union ; ∩ algorithms intersection

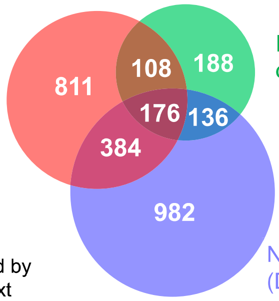

Fig 1. Number of hemorrhages detected by structured and free-text algorithms at the CHUV

Table 3. Validation of the algorithms (CHUV)

Information source	Recall (%)
ICD-10/CHOP codes	95
Laboratory data and prescription orders	60
Algorithms union	98
Algorithms intersection	52
ICD-10/CHOP codes ∩ Laboratory data and prescription orders	57

recall= true positives/(true positives+false negatives)

SwissMADE collaborators: M. Lutters, C. Lovis, P. Beeler, F. Rinaldi, PO Lang