Patient Safety Informatics
V. Koutkias et al. (Eds.)
IOS Press, 2011

doi:10.3233/978-1-60750-740-6-25

EU-ADR Healthcare Database Network *vs*. Spontaneous Reporting System Database: Preliminary Comparison of Signal Detection

Gianluca TRIFIRÒ [a, 1], Vaishali PATADIA [a,b], Martijn J. SCHUEMIE [a], Preciosa M. COLOMA [a], Rosa GINI [c], Ron HERINGS [d], Julia HIPPISLEY-COX [e], Giampiero MAZZAGLIA [f], Carlo GIAQUINTO [g], Lorenza SCOTTI [h], Lars PEDERSEN [i], Paul AVILLACH [j,k], Miriam C.J.M. STURKENBOOM [a] and Johan VAN DER LEI [a], on behalf of the EU-ADR group

[a] *Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands*
[b] *Astellas Pharmaceuticals*
[c] *Agenzia Regionale di Sanità della Toscana, Florence, Italy*
[d] *PHARMO Institute, Utrecht, The Netherlands*
[e] *University of Nottingham, UK*
[f] *Società Italiana di Medicina Generale, Florence, Italy*
[g] *Pedianet-Società Servizi Telematici SRL, Padova, Italy*
[h] *Department of Statistics, Università di Milano-Bicocca, Milan, Italy*
[i] *Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark*
[j] *Université Victor Segalen Bordeaux, France*
[k] *LERTIM, Faculté de Médecine, Université de la Méditerranée, Marseille, France*

Abstract. The EU-ADR project aims to exploit different European electronic healthcare records (EHR) databases for drug safety signal detection. In this paper we report the preliminary results concerning the comparison of signal detection between EU-ADR network and two spontaneous reporting databases, the Food and Drug Administration and World Health Organization databases. EU-ADR data sources consist of eight databases in four countries (Denmark, Italy, Netherlands, and United Kingdom) that are virtually linked through distributed data network. A custom-built software (Jerboa©) elaborates harmonized input data that are produced locally and generates aggregated data which are then stored in a central repository. Those data are subsequently analyzed through different statistics (i.e. Longitudinal Gamma Poisson Shrinker). As potential signals, all the drugs that are associated to six events of interest (bullous eruptions - BE, acute renal failure - ARF, acute myocardial infarction - AMI, anaphylactic shock - AS, rhabdomyolysis - RHABD, and upper gastrointestinal bleeding - UGIB) have been detected via different data mining techniques in the two systems. Subsequently a comparison concerning the number of drugs that could be investigated and the potential signals detected for each event in the spontaneous reporting systems (SRSs) and EU-ADR network was made. SRSs could explore, as potential signals, a larger number of drugs for the six events, in comparison to EU-ADR (range: 630-3,393 vs. 87-856), particularly for those events commonly thought to be

[1] Corresponding Author: g.trifiro@erasmusmc.nl

potentially drug-induced (i.e. BE: 3,393 vs. 228). The highest proportion of signals detected in SRSs was found for BE, ARF and AS, while for ARF, and UGIB in EU-ADR. In conclusion, it seems that EU-ADR longitudinal database network may complement traditional spontaneous reporting system for signal detection, especially for those adverse events that are frequent in general population and are not commonly thought to be drug-induced. The methodology for signal detection in EU-ADR is still under development and testing phase.

Keywords. Pharmacovigilance, electronic health records, drug safety, signal detection, spontaneous reporting database

Introduction

World Health Organization defines a drug safety signal as information on a possible causal relationship between an adverse event and a drug, which is unknown or incompletely documented [1]. Historically, spontaneous reporting systems (SRSs) for adverse drug reactions (ADRs) have been the cornerstone of signal detection in pharmacovigilance for the last four decades [2]. Cerivastatin and more recently rofecoxib stories highlighted the limitations of spontaneous reporting system with respect to the early detection of ADRs. The increasing availability of electronic healthcare records (EHRs) offers opportunities to investigate a wide spectrum of adverse drug effects and to detect signals closer to real time [3]. EHR databases present the additional advantage of large populations and long follow-up periods. A number of data mining techniques have been specifically developed for automatic detection of drug safety signals [2]. Currently, a number of ongoing international initiatives (SENTINEL [4], EU-ADR [5], PROTECT [6], and OMOP [7]) are aimed at testing the potential of signal detection using longitudinal electronic health record databases.

The EU-ADR (Exploring and Understanding Adverse Drug Reactions by integrative mining of clinical records and biomedical knowledge) project was funded by the European Commission and started in February 2008. The overall objective of the project was to design, develop, and validate a computerized integrative system that exploits data from EHRs and biomedical databases for the early detection of ADRs. Beyond the current state-of-the-art, EU-ADR led to the federation of different databases of EHRs, creating a resource of unprecedented size for drug safety monitoring in Europe (over 30 million patients from eight different databases). The initial stage of signal generation is followed by signal substantiation through causal reasoning, semantic mining of literature, and computational analysis of pharmacological and biological information, all with the aim of finding possible pathways that explain the drug-event associations.

As regard signal generation, in the EU-ADR project an event-based approach was adopted. A set of events warranting priority for monitoring in pharmacovigilance have been selected and inspected for their association with all possible drugs [8].

In this paper we describe the preliminary results of the comparison between EU-ADR healthcare network and two spontaneous reporting systems databases (Food and Drug Administration - Adverse Event Reporting System (FDA-AERS) and World Health Organization (WHO) Vigibase). As potential signal in the two systems, for the preliminary analyses we considered all the drugs being associated with the following six events that are deemed to be important in pharmacovigilance: Upper Gastrointestinal Bleeding (UGIB), Anaphylactic Shock (AS), Acute Myocardial

Infarction (AMI), Rhabdomyolysis (RHABD), Acute Renal Failure (ARF) and Bullous Eruption (BE).

1. Methods

1.1. Signal Detection in EU-ADR

The EU-ADR database network currently comprises of anonymised healthcare data from eight established European databases located in four countries: Health-Search (HSD, Italy). Integrated Primary Care Information (IPCI, Netherlands), Pedianet (Italy) and QResearch (United Kingdom) are general practice (GP) databases, while Aarhus University Hospital Database (Denmark), PHARMO (Netherlands), and the regional Italian databases of Lombardy and Tuscany are all comprehensive record-linkage systems in which drug dispensing data of a well-defined population is linked to a registry of hospital discharge diagnoses and other medical registries.

Due to the difference in coding schemes across various databases, the Unified Medical Language System (UMLS) was initially used as the terminology to define the events of interest [9]. Subsequently projection of the selected UMLS concept into different terminologies (i.e. READ, ICD9-CM, ICD10, and ICPC) was carried out.

In the EU-ADR project we adopted a distributed network approach that requires standardization of input files from the different databases. These input files (patient, drug, and event files) have been created locally by each database owner and have been subsequently elaborated through the purpose-built software called Jerboa© [10]. The software queries patient-level data in the different databases, which is later aggregated, and sent in encrypted format to a central repository for further analyses. For the analysis described in this paper, data from 1996 till 2010 has been contributed from six databases (QResearch and UNIMIB databases could not contribute data for this analysis). Several statistics were generated to detect all the associations between all the covered drugs and the six events of interest. Currently, the Longitudinal Gamma Poisson Shrinker (LGPS) posterior expectation of the incidence rate ratio higher than 2 and p-value<0.05 are the criteria that have been considered to distinguish between potential signals and non-signals [11]. The LGPS is a modification of the GPS method used in some spontaneous reporting system databases. These statistical approaches apply shrinkage to the frequentist estimates to reduce the chance of a false positive result. For the incidence rate ratios exposed time was compared with all non-exposed time including time exposed to other drugs. Based on empirically determined background incidence rates, for each event the minimum required amount of exposure was determined and the drugs not reaching this threshold were not tested as potential signals.

1.2. Signal Detection in FDA-AERS and WHO

Food and Drug Administration (FDA) - Adverse Event Reporting System (AERS) and World Health Organization (WHO) spontaneous reporting databases have been used as comparators. The FDA-AERS database is a computerized spontaneous reporting database that was established in 1969 to support the FDA's post-marketing safety

surveillance program and currently contains over 4 million reports of suspected adverse drug reactions (ADRs). FDA-AERS collects most of its reports from the USA.

The WHO spontaneous report database (Vigibase) was established in 1968 and is maintained by the Uppsala Monitoring Centre (UMC) [12]. VigiBase contains at the moment more than 4 million reports of suspected ADRs that are sent from the national centers of 95 countries participating in the WHO Programme for International Drug Monitoring.

Both databases collect reports from marketing authorization holders, healthcare professionals and consumers. Overlapping of the collected report in the two databases is present. The suspected adverse drug reactions are coded using the Medical Dictionary for Regulatory Activities (MedDRA). All the Preferred Terms (PTs) of MedDRA corresponding to the six events have been used.

As regard the drug coding, an internal mapping between the generic name and the ATC code has been created. A disproportionality analysis was performed using the above mentioned PTs and the drug-ATC mapping in FDA-AERS and WHO database from the beginning (1968-9) through the 3Q2010 data. Empirical Bayes Geometric Mean (EBGM) was used to detect signals. A threshold of EB05>2 (with number of reports>0) was applied, with EB05 being the lower band of 95% Confidence Interval of EBGM [13].

As preliminary comparison for signal detection in SRSs and EU-ADR, for each of the six events we calculated the number of drugs that could be investigated and we identified the potential signals. The number of drugs that can be investigated depends on the presence of at least one report of suspected ADR in spontaneous reporting databases and on the presence of at least one exposed case patient (i.e. patients exposed to the drug when the event occurred) in the EU-ADR database network.

Table 1. Overview of signal detection in FDA-AERS and EU-ADR for the six events under consideration.

	Spontaneous reporting databases				EU-ADR		
	FDA-AERS		WHO VigiBase				
Event	**N. of drugs that could be studied**	**Potential signals N (%)**	**N. of drugs that could be studied**	**Potential signals N (%)**	**N. of drugs that could be studied**	**Potential signals N (%)**	**Potential signals in both systems, N**
Acute myocardial infarction	791	38 (4.8)	630	37 (5.9)	856	143 (16.7)	6
Acute renal failure	2,626	354 (13.5)	3,002	302 (10.1)	461	171 (37.1)	40
Anaphylactic shock	1,443	144 (10.0)	2,679	269 (10.0)	265	47 (17.7)	13
Bullous eruption	2,053	289 (14.1)	3,393	225 (6.6)	228	42 (18.4)	13
Rhabdomyolysis	1,302	94 (7.2)	1,164	51 (4.4)	87	30 (34.5)	3
Upper GI bleeding	1,937	115 (5.9)	2,419	175 (7.2)	695	218 (31.4)	31

Legend: *N. of drugs that could be studied*=number of drugs that could be investigated as potential signals, which depends on the presence of at least one report of suspected adverse drug reactions in FDA-AERS and on at least one exposed case patient in EU-ADR. *Potential signal:* statistically significant association between drug and event, based on specific analyses as described in paragraphs 1.1 and 1.2.

2. Results

Table 1 shows for each event the number of drugs that could be tested as potential signals and the number of signals being detected in the two spontaneous reporting databases and the EU-ADR system. The unit of analysis for signals is represented by single drug-event association. Overall, spontaneous reporting systems could explore, as potential signals, a larger number of drugs in association with the six events under study, in comparison to EU-ADR (range: 630-3,393 vs. 87-856). This difference was even higher for the events that are thought to be potentially drug-induced (i.e. BE: 2,053 in FDA and 3,393 in WHO vs. 228 in EU-ADR; ARF: 2,626 in FDA and 3,002 in WHO vs. 461 in EU-ADR). On the contrary, concerning the analysis for AMI a larger number of drugs could be investigated in EU-ADR (856) than SRSs (791 in FDA and 630 in WHO).

Overall, higher proportion of potential signals is detected in EU-ADR as compared to SRSs (17-37% vs. 5-14%). For the signal generation new methodologies are currently under development in EU-ADR.

The potential for signal detection in both EU-ADR and spontaneous reporting systems varies across events. The highest proportion of signals detected in SRSs was reported for BE, ARF and AS, while for ARF, UGIB and RHABD (for this event however a very low number of drugs could be tested) in EU-ADR.

3. Conclusion

The potential of EU-ADR database network for drug safety signal detection is promising particularly for those adverse events that have high frequency (i.e. acute myocardial infarction) in general population. Data mining of longitudinal electronic medical records may particularly complement traditional analyses on spontaneous reporting systems in the signal detection, especially for those frequent adverse events that are not traditionally thought to be drug induced. The implementation of additional analyses in the EU-ADR system is still ongoing. In the final EU-ADR system, a panel of statistical analyses will allow a greater precision of signal detection. In addition, automatic search in the scientific literature and summary of product characteristics will filter out the already known signals among those being initially identified in EU-ADR. On the other hand, signals will be substantiated by a computer-assisted exploration of biological plausibility in the context of current biomedical knowledge to reduce the false positive signals.

Acknowledgments

This research has been funded by the European Commission Seventh Framework Programme (FP7/2007-2013) under grant no. 215847 – The EU-ADR Project.

The authors also wish to thank the NLM for making UMLS available free of charge.

References

[1] M. Stahl et al., Assessing the impact of drug safety signals from the WHO database presented in "SIGNAL": results from a questionnaire of National Pharmacovigilance Centres, *Drug Saf* **26** (2003), 721–7.
[2] E.M. Rodriguez, J.A. Staffa, D.J. Graham, The role of databases in drug postmarketing surveillance, *Pharmacoepidemiol Drug Saf* **10** (2001), 407–10.
[3] S. Schneeweiss, J. Avorn, A review of uses of health care utilization databases for epidemiologic research on therapeutics, *J Clin Epidemiology* **58** (2005), 323–37.
[4] http://www.fda.gov/Safety/FDAsSentinelInitiative/ucm2007250.htm, last access on March 2011.
[5] G. Trifirò et al., The EU-ADR project: preliminary results and perspective, *Stud Health Technol Inform* **148** (2009), 43–9.
[6] Pharmacoepidemiological Research on Outcomes of Therapeutics by a European Consortium – PROTECT project. Available at: http://www.imi-protect.eu/, last access on March 2011.
[7] P.E. Stang et al., Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership, *Ann Intern Med* **153** (2010), 600–6.
[8] G. Trifirò et al., Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor?, *Pharmacoepidemiol Drug Saf* **18** (2009), 1176–84.
[9] P. Avillach et al., Design and evaluation of a semantic approach for the homogeneous identification of events in eight patient databases: a contribution to the European EU-ADR project. *Stud Health Technol Inform* **160** (2010), 1085–9.
[10] P.M. Coloma et al., Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project, *Pharmacoepidemiol Drug Saf* **20** (2011), 1–11.
[11] M.J. Schuemie, Methods for drug safety signal detection in longitudinal observational databases: LGPS and LEOPARD, *Pharmacoepidemiol Drug Saf* 2010 Oct 13 [Epub ahead of print].
[12] M. Lindquist, The WHO Adverse Reaction Database: Basic Facts. In: http://www.who-umc.org/graphics/4789.pdf. Uppsala: Uppsala Monitoring Centre; 2004, last access on March 2011.
[13] A. Szarfman, S.G. Machado, R.T. O'Neill, Use of screening algorithms and computer systems to efficiently signal higher-than-expected associations of drugs and events in the US FDA's spontaneous reports database, *Drug Safety* **25** (2002), 381–92.