The EU-ADR Web Platform: delivering advanced pharmacovigilance tools

José Luis Oliveira¹*, Pedro Lopes¹, Tiago Nunes¹, David Campos¹, Scott Boyer², Ernst Ahlberg², Erik M. van Mulligen³, Jan A. Kors³, Bharat Singh⁴, Laura I. Furlong⁴, Ferran Sanz⁴, Anna Bauer-Mehren⁴, Maria C. Carrascosa⁴, Jordi Mestres⁴, Paul Avillach⁵,⁶, Gayo Diallo⁶, Carlos Díaz Acedo⁷ and Johan van der Lei³

¹DETI/IEETA, University of Aveiro, Aveiro, Portugal
²AstraZeneca, Molndal, Sweden
³Erasmus University Medical Center, Rotterdam, The Netherlands
⁴Research Programme on Biomedical Informatics (GRIB), IMIM Hospital del Mar Research Institute and Universitat Pompeu Fabra, Barcelona, Catalonia, Spain
⁵LERTIM, EA 3283, Faculté de Médecine, Université de Aix-Marseille, Marseille, France
⁶LESIM-ISPED, Université de Bordeaux, Bordeaux, France
⁷Synapse Research Management Partners, Barcelona, Spain

ABSTRACT

Purpose Pharmacovigilance methods have advanced greatly during the last decades, making post-market drug assessment an essential drug evaluation component. These methods mainly rely on the use of spontaneous reporting systems and health information databases to collect expertise from huge amounts of real-world reports. The EU-ADR Web Platform was built to further facilitate accessing, monitoring and exploring these data, enabling an in-depth analysis of adverse drug reactions risks.

Methods The EU-ADR Web Platform exploits the wealth of data collected within a large-scale European initiative, the EU-ADR project. Millions of electronic health records, provided by national health agencies, are mined for specific drug events, which are correlated with literature, protein and pathway data, resulting in a rich drug–event dataset. Next, advanced distributed computing methods are tailored to coordinate the execution of data-mining and statistical analysis tasks. This permits obtaining a ranked drug–event list, removing spurious entries and highlighting relationships with high risk potential.

Results The EU-ADR Web Platform is an open workspace for the integrated analysis of pharmacovigilance datasets. Using this software, researchers can access a variety of tools provided by distinct partners in a single centralized environment. Besides performing standalone drug–event assessments, they can also control the pipeline for an improved batch analysis of custom datasets. Drug–event pairs can be substantiated and statistically analysed within the platform’s innovative working environment.

Conclusions A pioneering workspace that helps in explaining the biological path of adverse drug reactions was developed within the EU-ADR project consortium. This tool, targeted at the pharmacovigilance community, is available online at https://bioinformatics.ua.pt/euadr/. Copyright © 2012 John Wiley & Sons, Ltd.

Key words—pharmacoepidemiology; pharmacovigilance; adverse drug reactions; drug safety; in silico pharmacology

INTRODUCTION

Contemporary prevention and treatment of diseases revolves around a dynamic medication market where pharmaceutical companies compete, aiming to investigate, develop and introduce new drugs in daily healthcare provision. Despite the expected therapeutic benefit of these innovations, drug safety is a major concern for regulatory health authorities, drug companies and researchers as several marketed drugs continue to pose serious risks to the well-being of many patients, having become, in recent years, one of the leading causes of mortality.¹

The traditional approach tackles this problem from a pre-market perspective, that is, conditional drug approval. Both the European Medicines Agency² and the US Food and Drug Administration³ establish rigorous guidelines for new medicine approval, requiring intense testing and trials, which result in a long and complex lab-to-market development cycle.⁴ Along with these guidelines,
pharmaceutical companies must also define thorough risk management plans for post-market drug stages.\textsuperscript{5,6}

Consequently, the relevance of post-market pharmacovigilance in the health domain has been growing steadily over the last 4 decades.\textsuperscript{7,8} Research in this area involves the exploration and assessment of signals, defined by the World Health Organization as undisclosed assertions on direct relationships between adverse events and a drug.\textsuperscript{9} Clinicians use spontaneous reporting systems (SRSs) to identify adverse drug reactions. These systems empower physicians with tools to report suspicions on certain drugs to a pharmacovigilance centre. Latest advances take these tools even further, completing the drug loop by providing a reporting infrastructure to pharmacists, patients and litigation lawyers, another group of adverse drug reaction (ADR) reporters, which in certain cases (particularly in the USA) may amount to 90% of the electronic ADR records causing a certain bias. Although many ADRs were detected through these systems, there are inherent limitations that hamper signal detection.\textsuperscript{10,11} They depend entirely on the ability to recognize an adverse event as being related to the drug and on his availability to report the case to the local spontaneous reporting database. The greatest limitations, therefore, are underreporting and biases due to selective reporting. Investigations have shown that the percentage of ADRs being reported varies between 1% and 10%.\textsuperscript{12–14}

Consequently, there is a high demand for novel software tools capable of improving the post-marketing drug monitoring workflow.\textsuperscript{15} Moreover, it is of utmost importance to explain the plausibility of ADR at the light of current scientific knowledge. By exploring modern knowledge engineering technologies, we are able to overcome the limitations associated with insufficient clinical trial data, complex monitoring statistics and closed general practice data silos. Text and data-mining tools, combined with distributed computing strategies, pave the way for better in silico signal identification and ADR assessment.\textsuperscript{16}

Adverse drug reaction reporting and analysis

Hårmark and Grooteest research explains the underlying pharmacovigilance concerns with current drug evaluation approaches.\textsuperscript{17} Although drug safety concerns are becoming more prominent, the lack of adequate software to correctly understand ADRs continues to challenge the pharmaceutical industry and research community.\textsuperscript{18,19}

The risk associated with any marketed drug triggers critical safety concerns, which, in their turn, leverage a constant revision and update of medical products’ information. For these tasks, modern ADR monitoring becomes essential. Despite the complex set of drug trials, including the final randomized double blind evaluation, clinical trials data are in most scenarios insufficient to assess drug risk in a larger or particularly susceptible population. Rare ADRs, ADRs identified in particular population cohorts or ADRs with long latency, require intensive post-marketing drug analysis.

At this level, SRSs come to play. Pharmacovigilance centres task is to collect these reports, generating enough data to inform stakeholders of potential risks as soon as they appear in the system. Despite the invaluable data coming from SRS, their data alone is meaningless without further processing and additional information. Viewing SRS as independent entities makes it nigh impossible to establish direct relationships between the causes (a drug or drug interaction) and consequences (a phenotype). Hence, to extract meaningful insights from these SRS records, we need to rely on advanced data-mining techniques.\textsuperscript{20} These provide distinct perspectives over acquired data and their connections to other information topics.\textsuperscript{21}

Another strategy is in place to complement SRSs. Intensive monitoring systems rely on prescription data, forcing drug prescribers to ask about any adverse reaction during the drug intake cycle. Once these data are collected, they are processed for signal evaluation. Unlike SRS, which is based on monitoring specific drugs over a controlled period, intensive reporting relies on a non-interventional observational cohort. Hence, generated data are much nearer real-world scenarios than data obtained through SRS. Intensive reporting also renewed the interest in the importance of health information systems (HISs) and general practice research databases.

The EU-ADR initiative

Despite the myriad of international developments in these fronts, most efforts approach this problem from a pre-market approach, focusing on conditional drug approval and defining guidelines for risk management plans. Hence, modern projects, such as EU-ADR\textsuperscript{22} or RADAR,\textsuperscript{23} define a proactive strategy for post-marketing drug assessment. To overcome the ‘reporting bias’ and under-reporting of physicians, the EU-ADR solution was based on automatically exploiting the data stored in large electronic health record (EHR) databases. Modern regional and national HISs tend to store miscellaneous information regarding patients’ clinical history, including drug prescriptions, vaccinations, height, weight or laboratory test results, among others.\textsuperscript{24} These wide collections of data are traditionally a good general representation of region demographics. Furthermore,
collected data are already used for pharmacoepidemiology, disease epidemiology and, to a lesser extent, pharmacoeconomics including drug usage monitoring.25–27 From a pharmacovigilance perspective and in a European or worldwide scale, mining the amount and type of data collected in these databases is of tremendous importance for an improved post-marketing drug evaluation.28

The foundation for this strategy is doing in-depth semantic data mining on the wealth of EHRs to generate filtered data that can be easily substantiated through distributed computational tools.16,29 The final output, a ranked signal list, provides a broad look over identified signals and their significance in health risk.

The EU-ADR Web Platform tackles these challenges, extending the availability of existing tools to every stakeholder, through a web-based pharmacovigilance suite. This system enables an insightful exploration of pharmacovigilance signals’ evolution resulting in a comprehensive risk evaluation. This is possible through innovative features such as the creation of custom drug studies, the remote execution of signal substantiation workflows or the cross-analysis against millions of anonymous EHRs.30 Before starting, we recommend the consultation of the user guide, available at http://bioinformatics.ua.pt/euadr/public/guide.pdf.

METHODS

The European EU-ADR Project exploits partner data from national EHRs and HISs of about 30 million European patients,24 channelling it through state-of-the-art distributed computing software and enriching signal detection.31 This large-scale drug safety monitoring relies in various mining, epidemiological, statistical and computing techniques to assess acquired data and generate a ranked signal list—Figure 1.

With EU-ADR’s huge knowledge base in place, innovative methods to access and explore collected data are required because many ADRs can be biologically explained if we are able to integrate current biomedical knowledge. We call this process signal substantiation.16 This signal substantiation is performed through several distributed software, streamlined into a single computational workflow.32 On this bioinformatics context, the Taverna workbench arises as the de facto platform for workflow creation, management and execution.33

The Medline ADR signal filtering workflow automates literature analysis tasks by assessing a list of publications regarding a specific signal. The algorithm adopts a semantics-based approach that processes Medline annotations looking for particular MeSH terms.34 This workflow’s output is a direct relationship between an ADR and its descriptions in Medline, if present.

Second, the signal filtering co-occurrence process is divided into three similar workflows, each targeting a distinct resource. These evaluate the relationships between drugs and side effects that might have been reported previously in Medline literature (Medline Co-occurrence) or drug databases such as DailyMed35 or DrugBank.36 These workflows use statistical and text-mining techniques to evaluate drug names, Anatomical Therapeutic Chemical Classification System (ATC) codes and event co-occurrences in the indexed resources.

At last, the Signal Substantiation workflow analyses the drugs, proteins and pathways interaction graphs. This involves searching for proteins targeted by the drug and associated with the clinical event, directly or through biological pathways. The algorithm generates drug-target and event-protein profiles that are searched for common sets of proteins, the intersecting portion of the graph.

Figure 1. EU-ADR data flow. (1) Data from electronic health record (EHR) resources is semantically harmonized for data extraction. (2) Extracted data is mined for drug–event pairs and other relationships. (3) The signal generation process takes mined data and forms the first ranked signal dataset. (4) The signal substantiation process re-ranks the signal list, based on evidences from biomedical databases and literature, in silico simulations and pathway analyses. (5) The EU-ADR Web Platform enables completing the retrospective and prospective system validation.
These five workflows accept a similar input, a drug–event pair, and produce a similar output, standardized XML. Workflow interactions are made possible by EU-ADR’s XML schema language (http://bioinformatics.ua.pt/euadr/euadr_types.xsd). The data flow from and to workflows is exchanged in XML described using an EU-ADR internal schema. This is a true interoperability enabler as data are shared in a format that is understood by all tools in the EU-ADR ecosystem.

Despite enabling an independent use within the mentioned Taverna workbench, EU-ADR workflows may still be out of reach of the general pharmacovigilance community with low computer expertise. Further, combining these workflows’ results is an essential step towards a better understanding of drug–event relationships. As such, to foster an easier use and promote the aggregation of results, a centralized workflow management and execution tool is needed. This complements local workflow use for individual analysis with remote workflow execution for processing large heterogeneous datasets. Moreover, when executed online in the EU-ADR Web Platform, workflow results are displayed in a specialized interface, designed to highlight their relevant parts and facilitate evidence analysis. This unique interface contrasts with the raw text and XML data obtained directly from Taverna.

Evidence combination is a central part of signal substantiation. Although each workflow’s result has a value on its own, through combination of different results, we can leverage knowledge from multiple sources and better assess the plausibility of a given drug–event relationship. Each EU-ADR workflow yields a binary score, representative of evidence for a given drug–event relationship being found or not. Then, using Dempster–Shafer theory, we combine evidence from several disparate sources and arrive at a degree of belief that takes into account all the available evidence—Figure 2. In summary, the Dempster–Shafer algorithm will evaluate the initial data combined with workflow results to reach a measurable belief level that a particular drug–event pair has a low, medium or high risk.

In a heterogeneous ecosystem offering different means to evaluate signal plausibility, it is important to weigh the trustworthiness of one method against another. Hence, for greater flexibility in signal detection, we must customize the reliability of individual substantiation methods, both nominal workflow’ scores and numerical values obtained from statistical analysis of EHR data. Because confidence in any given substantiation method is highly subjective, users should be able to tailor the evidence combination process for their needs and save their settings privately on the system.

The EU-ADR Web Platform also tackles the data-sharing and research reproducibility issues. By

![Evidence combination process](Figure 2. Evidence combination process. Various evidence scores from multiple sources are combined into a single score using configurable reliability and accuracy parameters for each evidence source. The Dempster–Shafer theory is used to arrive at a degree of belief that takes into account all the available evidence and facilitate detection of possible adverse drug reactions. EHR, electronic health record; ADR, adverse drug reaction)
storing data and workflows online, the EU-ADR Web Platform enables replicating research strategies to follow previous procedures, to confirm previous results or to test if there are novel substantiation outcomes. As the same data and services are used, researchers are assured that their results are unique and longstanding. Data-sharing environments can also be enabled through collaborative groups (projects), unlocking read and write access. Additionally, existing datasets can be shared, in an *ad hoc* fashion, to any number of users.

To build this complex system while maintaining focus on the core features that make it unique implied the implementation of commonly required features through delegate third-party frameworks and libraries. Hence, the EU-ADR Web Platform is a collaborative workspace built over a solid foundation of open-source software components. Users interact directly with the platform client, a highly responsive application that runs inside any modern browser.

Because web-based distributed systems are affected by connection quality and inherently prone to availability issues, the platform depends on remote resources only for data submission, data loading and signal substantiation. This means once a dataset and related evidence are loaded, connectivity loss does not hamper system usage. Moreover, all unexpected errors are reported and logged to the server whenever possible, effectively leading to continuous improvement of the system over time.

RESULTS

Setup

The EU-ADR Web Platform is sustained by a distributed computerized system combining multiple components in a single software ecosystem. Figure 3 highlights the data flow from the user submissions to the multiple workflow interactions.

EU-ADR workflows play an active role in the EU-ADR Web Platform, as they are required for data analysis and signal evaluation. The challenging tasks of accessing and executing workflows required the development of a new workflow execution engine, enabling real-time web-based communication with the workflows.

Because Taverna is in charge of workflow execution, we need to feed the services with input data, manipulate intermediate results and extract the resulting output documents. The final data is then parsed by the web platform and presented to users on the client side in a way that facilitates evidence analysis and assessment. A thin wrapper was developed in Java, launching parameterized calls to the Taverna command-line tool, which runs in its own process, controlled by system calls.

Workflow execution is a non-blocking asynchronous process. From a usability perspective, this results in a more interactive experience as the workspace can still be used during background workflow execution. Furthermore, EU-ADR’s workflows involve services that are not physically or logically co-located, leveraging a truly distributed service execution.

The client application uses a myriad of advanced user interaction components to provide a unique perspective on the huge drug datasets and easy access to data exploration features. Investigation of any drug–event pair does not end after the primary relative risk assessment as evidence can be combined to reach a final score, helping the separation between spurious signals and potential ADRs.

**Feature highlights**

The EU-ADR Web Platform is built to support advanced pharmacovigilance studies. The invite-based...
registration system allows authorized researchers to join the web platform by giving them access to a personal closed workspace. Registered users are able to upload and analyse drug–event datasets, create targeted drug studies, collaborate with their research peers through the available sharing features and execute EU-ADR workflows locally or remotely.

EU-ADR Web Platform features are available in an online user portal, divided in Datasets and Workflow views. These sections provide an entry point for exploring drug–event data and accessing project workflows, respectively.

The Dataset list view, shown in Figure 4, enables managing each user’s datasets. Datasets are divided in two sections, My Datasets, listing the user personal datasets, and Shared by Others, listing datasets shared with the user. Both sections include a dataset management action box, allowing the upload of new standardized datasets or the creation of drug-specific datasets, among others. Members of the EU-ADR project have access to an additional section, the EU-ADR Project collaborative workspace. This secure workspace facilitates cooperative study of EU-ADR datasets amid project members and assures the confidentiality of all project-private data.

Drug–event datasets can be imported to the system from plain-text files in CSV or TXT format or Microsoft Excel spreadsheets in XLS or XLSX format. Each imported file can include up to 5000 drug–event pairs in a standardized format, where the mandatory fields are the drug ATC code and the EU-ADR event acronym—Table 1. Apart from an optional ‘Name’ field, treated as the drug name, each signal can contain any number of additional attributes, which are imported to the platform database and can later be visualized in the dataset view.

Targeted datasets are focused on a single drug, statistically analysed against the 30 million anonymous EU-ADR records. The dataset signal list is automatically generated from all signals in the database. That is, the drug is related to EU-ADR events covering 11 clinically relevant adverse reactions.

Double clicking on a dataset loads its content in a new workspace tab. This view lists all dataset signals and their respective data in a single table. This listing is enriched when the substantiation process is triggered (Substantiate action button), filling in the results from each external workflow and from the evidence combination analysis.

The Workflows menu loads the five EU-ADR workflows. In this view, each workflow is described and a variety of actions are displayed. Workflows can be exported for local execution or substantiated remotely with custom relationships or using the example signals.

The combination of dataset management features with targeted drug–event analysis features delivers an innovative framework for filtering and substantiation. With the inclusion of direct sharing possibilities, the EU-ADR Web Platform enables the creation of a pharmacovigilance collaborative research environment.

---

**Figure 4.** EU-ADR Web Platform dataset list view. (1) Access to workflows list view. (2) ‘My Datasets’ section listing user’s dataset list. (3) Dataset action buttons, from left to right: create new targeted datasets, import dataset from local file, export online dataset to local file, open selected dataset, share selected dataset, delete selected dataset, help. (4) Dataset list table detailing dataset name, description, number of drug–event pairs and creation date. (5) Access to ‘Shared by Others’ section, listing datasets shared with the user.
DISCUSSION

For an assessment of EU-ADR Web Platform’s applicability to a real-world research workflow, a sample drug analysis scenario is presented here. A researcher interested in studying potential adverse reactions of patients treated with a given drug, Drug_X for the purpose of this discussion, begins its study by automatically generating a dataset focused on the drug under investigation. The system then combines this drug with the 11 potential adverse events considered in EU-ADR, substantiates the resulting signals using the available workflows and combines all individual pieces of evidence into an aggregate score representing the predicted risk of each drug–event relationship—Figure 5. Signals classified as moderately or highly risky should be further investigated by analysing presented evidence and following hyperlinks to biomedical literature, as well as to external drug and biological data resources. A more detailed use case is available in the user guide (http://bioinformatics.ua.pt/euadr/public/guide.pdf).

Pharmacovigilance research over the last decades has focused mainly on evaluating the best strategies to identify and measure specific ADRs in a post-marketing stage.40–42 The EU-ADR initiative further expands this trend by introducing a complete framework for drug–event interaction analysis, from EHRs data sources to a researcher-oriented web-based workspace.

To our knowledge, the EU-ADR Web Platform is the only current tool allowing researchers to exploit the wealth of data for a vast European-wide cohort. It enables independent drug analysis crossed against the millions of collected records. Furthermore, rather than

| ATC   | Name     | EventType | Exposure | Events | RR(MH) | ...
|-------|----------|-----------|----------|--------|--------|-------
| A01AA01 | Sodium fluoride | A1I | 5,302,087 | 4 | 3,18 | ...
| A01AA02 | Sodium fluoride | AMI | 4,897,540 | 6 | 2,39 | ...

Table 1. EU-ADR standard dataset format. EventType and ATC fields are the mandatory attributes that make up a potential ADR signal. The recommended Name field represents the drug name. If omitted, names of drugs are looked up in an internal drug database. Any additional signal attribute is imported as is and later presented in the application’s dataset visualization interface.

Figure 5. EU-ADR Web Platform results for an undisclosed drug (Drug_X) exploration scenario containing the signal list that results from distributed workflow outputs and evidence combination statistical analysis. Workflow results are labelled with Y in case sufficient evidence is found to support a potential drug–event relationship or N otherwise. Evidence combination yields a score of H, M or L, indicating high, moderate or low risk, respectively, of a drug–event relationship being in fact an adverse drug reaction signal.

being a single proof-of-concept application for the EU-ADR research project, this platform opens the door for broader ADR assessments beyond the limited EU-ADR event scope.

CONCLUSION

The EU-ADR European project embraces innovative pharmacovigilance research methodologies through the creation of a web platform providing advanced drug data exploration and assessment features. Whereas, in the past, post-marketing drug assessment required intense validation tasks, the in silico pharmacology community is now endowed with the tools required to quickly analyse specific ADRs, further improving drug safety monitoring.

The EU-ADR Web Platform enables streamlined access to drug dataset analysis features, including the evaluation of results from EU-ADR workflows and the sharing of data amongst research partners, all within a highly responsive and unique web-based workspace, which is available at http://bioinformatics.ua.pt/euadr.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

KEY POINTS

- Progress in pharmacovigilance demands new methods to further improve data exploration from traditional SRSs. Advanced tools are in place to mine data from general practitioners research databases, establishing useful connections to other well-known resources.
- Web services for the analysis of drug–event associations were developed, requiring the implementation of service composition strategies to foster interoperability within the pharmacovigilance software ecosystem.
- A unique web-based workspace, the EU-ADR Web Platform, is introduced to deliver advanced pharmacovigilance software to everyone, empowering the research community with pioneering tools to identify, monitor and evaluate ADRs.

ACKNOWLEDGEMENTS

We wish to thank all the members of the EU-ADR project. This work was supported by the European Commission (EU-ADR, ICT-215847), FCT (PTDC/ EIA-CICO/100541/2008) and Instituto de Salud Carlos III FEDER (CP10/00524). The Research Programme on Biomedical Informatics (GRIB) is a node of the Spanish National Institute of Bioinformatics (INB).

REFERENCES


