MEDINFO 2010
C. Safran et al. (Eds.)
IOS Press, 2010

doi:10.3233/978-1-60750-588-4-518

Documentation in Pharmacovigilance: Using an ontology to extend and normalize Pubmed queries

Denis Delamarre[a], Agnès Lillo-Le Louët[b], Laetitia Guillot[a], Anne Jamet[b], Eric Sadou[c], Theo Ouazine[a], Anita Burgun[a], Marie-Christine Jaulent[c]

[a]INSERM U936, Université Rennes 1, IFR 140, Rennes, France,
[b]Centre Régional de Pharmacovigilance, Hôpital Européen Georges Pompidou, Assistance Publique-Hôpitaux de Paris, Paris, France
[c]INSERM UMR_S 872, Eq. 20, Paris, F-75006 France; Université René Descartes, Paris, F-75006 France

Abstract

Objectives: To assess and understand adverse drug reactions (ADRs), a systematic review of reference databases like Pubmed is a necessary and mandatory step in Pharmacovigilance. In order to assist pharmacovigilance team with a computerized tool, we performed a comparative study of 4 different approaches to query Pubmed through ADR-drug terms. The aim of this study is to assess how an ontology of adverse effects, used to normalize and extend queries, could improve this search. Material and Method: The ontological resource OntoEIM contains 58,000 classes and integrates MedDRA terminology. The entry point is a ADR-Drug term and the four methods are (i) a direct search on Pubmed (ii) a search with a normalized query enhanced with domain-specific Mesh Heading criteria, (iii) a search with the same elaborated query extended to the MeSH sub-hierarchy of the adverse effect entry and (iv) a search with a set of MedDRA terms grouped by subsomption in the OntoEIM ontology. For each of the 16 queries performed and analysed, relevant publications are selected "manually" by two pharmacovigilant experts. Results: The recall is respectively of 63%, 50%, 67% and 74%, the precision of 13%, 26%, 29% and 4%. The best recall is provided by the ontology-based method, for 4 cases out of 16 this method returns relevant publications when the others return no results. Conclusion: Results show that an ontology-based search tool improves the recall performance, but other tools and methods are needed to raise the precision.

Keywords:

Adverse drug reaction reporting, Adverse drug reaction, Information retrieval, Databases bibliographic, PubMed, Ontology.

Introduction

Pharmacovigilance is the science focusing on detection, analysis and prevention of adverse drug reactions (ADRs). Spontaneous reports of ADRs by health care professionals allows the collection of case reports by central agencies and/or pharmaceutical companies [1]. Case reports (namely *ADR*-Drug pairs) are coded with the WHO-ART[1] and/or MedDRA[2] terminologies for the ADR and the ATC terminology for the drug, and stored in databases that constitute putative knowledge on suspected adverse drug reactions.

The pharmacovigilance team has the task to code newly reported ADR-drug cases but also to document and analyze them with relevant and up-to-date information retrieved from various sources. The sources presently requested during this review task are 1) drug summary of the product's characteristics (SPC), 2) pharmacovigilance-related data sources (such as Martindale, Meyler's Side effects of Drugs, other databases like Micromedex), 3) medical literature 4) previously reported cases. We address the documentation issue in the context of the Vigitermes project[3] which aims at developing a semantic portal to improve documentation of pharmacovigilance case reports for the pharmaceutical industry and regulatory authorities [2]. In the Vigitermes platform, for a given case, one action button performs the search on the drug SPC, Pubmed and pharmacovigilance databases. The PharmARTS tool is used to group cases with a close meaning, it relies on an ontology of adverse drug reactions (OntoEIM) that is based on the formal definitions of WHO-ART and MedDRa [3, 4]. In the present work, our objective is to evaluate the benefit of using the OntoEIM ontology in the request process on PubMed database. Our aim is to automate the search of relevant publications that correspond to a given ADR case. In a previous work, we presented a Pubmed querying web service for retrieving abstracts from the MEDLINE database based on Mesh heading and Mesh qualifier pattern criteria [5]. In this study, we performed a comparison of four different search methods using our Pubmed querying web service.

First, terms clustering and information retrieval with MEDLINE are presented in the background section. We then

[1] World Health Organization - Adverse Reaction Terminology
[2] Medical Dictionary for Drug Regulatory Activities
[3] http://vigitermes.univ-rennes1.fr

present the OntoEIM ontology and PharmARTS, the browsing tool used to query the ontology. The general architecture of the Vigipubmed toolbox is described and we detail the comparison study used to assess the "plus-value" of the ontology. Results are reported on a set of 13 ADR-drug cases are presented and discussed.

Background

Information retrieval with Medline

Several studies have shown that researchers report difficulties when searching electronic resources to document adverse drug reactions [6, 7]. Among others, these difficulties are due to poor indexing, to the wide variety of articles that may potentially be useful, to variation in their qualities, as well as the lack of tools to perform systematic search. Literature, specifically MEDLINE, is the main source of documentation used to detect whether a drug may be responsible for ADRs. Nowadays, literature searches are mainly done manually. However, the exhaustive retrieval of such information may not be straightforward, for three reasons. Firstly, the query is not easy to formulate: the terms have to be translated in English (for non-English speakers), and even mapped to MeSH entry terms. Secondly, the large number of publications in MEDLINE makes the search difficult and time-consuming. Thirdly, it is not easy to determine manually the level of evidence that should be the most appropriate to characterize the relation between a drug and a possible adverse effect.

Grouping similar ADRs

ADRs are coded with MedDRA or WHO-ART in pharmacovigilance databases. The MedDRA terminology, recommended for the description of ADRs by the pharmaceutical industry and regulatory authorities, includes entirely the WHO-ART terminology, but is considerably more extensive (15,000 preferred terms versus only 3500 for WHO-ART).

The structure of relationships used to organize terms is particularly important to retrieve similar medical conditions in the database and has a direct impact on the specificity and sensitivity of pharmacovigilance signal detection [8]. We have previously shown that neither WHO-ART nor MedDRA allows similar clinical conditions to be clustered together due to the lack of polyhierarchy. For instance in MedDRA, the term "gastric ulcer hemorrhage" is linked to the "gastric ulcer and perforations" term but not to the "gastric and esophageal hemorrhage" term. This deficiency limits the detection and the evaluation of ADRs [9].

We have developed an ontology (OntoEIM) describing ADRs that enhances the structural organisation of terms [3]. A web tool, PharmARTS, was developed as an interface to this ontology [10]. This ontology is under evaluation in the context of signal detection [4]. We investigate in the present article the relevance of this ontological resource to retrieve relevant articles in Medline for a given query.

Material and Method

Material

The OntoEIM iontology was previously developed and mapped with MedDRA and WHO-ART terminologies [3] PharmARTS, an online web service tool developed to query the ontology [10], is used to extend the query by grouping .MedDRA and WHO-ART terms with close meaning using OntoEIM.

OntoEIM

OntoEIM includes ADRs concepts obtained by aligning WHO-ART and MedDRA terms with SnomedCT[4], using the synonymy link in the metathesaurus of UMLS[5]. Relations in the ontology are associative relationships extracted from SnomedCT (e.g. "bladder neoplasm" is associated with the localization "bladder structure") and the relationship "is a" is used to indicate taxonomic relationships (e.g. "renal failure" is a "renal disease"). The ontology contains 5 798 WHO-ART classes, 37,892 MedDRA classes and 14,342 SnomedCT classes and includes 1,621 defined classes. The primary concepts of the hierarchy account for 4.6% of SnomedCT (308,677 concepts in the used version).

PharmARTS

PharmARTS is a web service developed in JAVA [10]. This tool can be used for querying (related to a clinical condition) and for grouping terms together by subsumption, (e.g. the terms "BLOOD_TRIGLYCERIDES_ABNORMAL" and "CHOLESTEROL_BLOOD_EXCESSIVE" could be used to encode two cases of dyslipidemia). PharmARTS can also be used to provide a more effective visual display of groups of terms and of associated pharmacovigilance cases [10].

Method

In order to assess the contribution of the ontology in the retrieval of relevant documentation from MEDLINE, we conducted a comparative study of four methods used to query the Pubmed database. Figure 1 displays the web service workflow of the evaluation system. The entry point is a pair of ADR-drug terms of a given case.

[4] http://www.ihtsdo.org/
[5] http://www.nlm.nih.gov/research/umls/

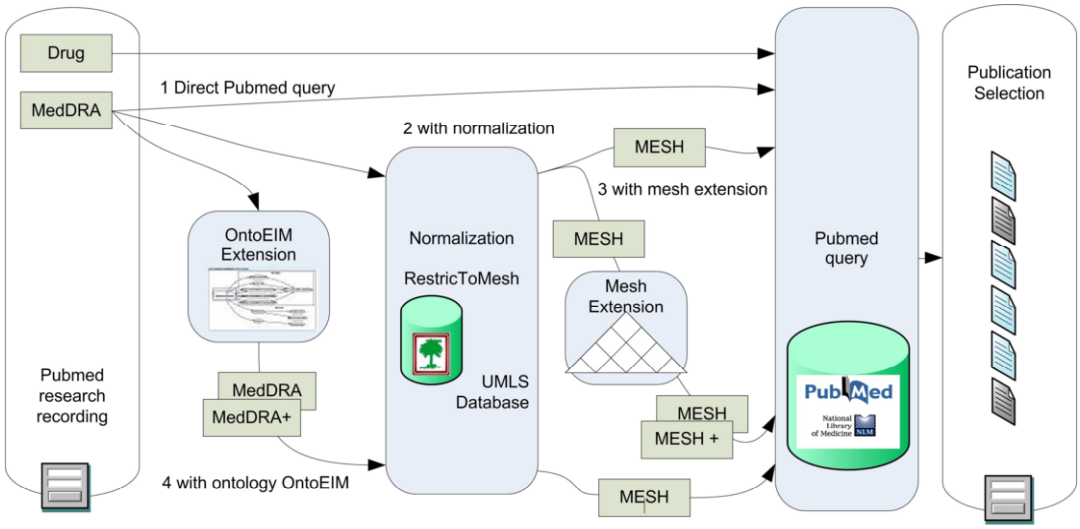


Figure 1- Web services work flow

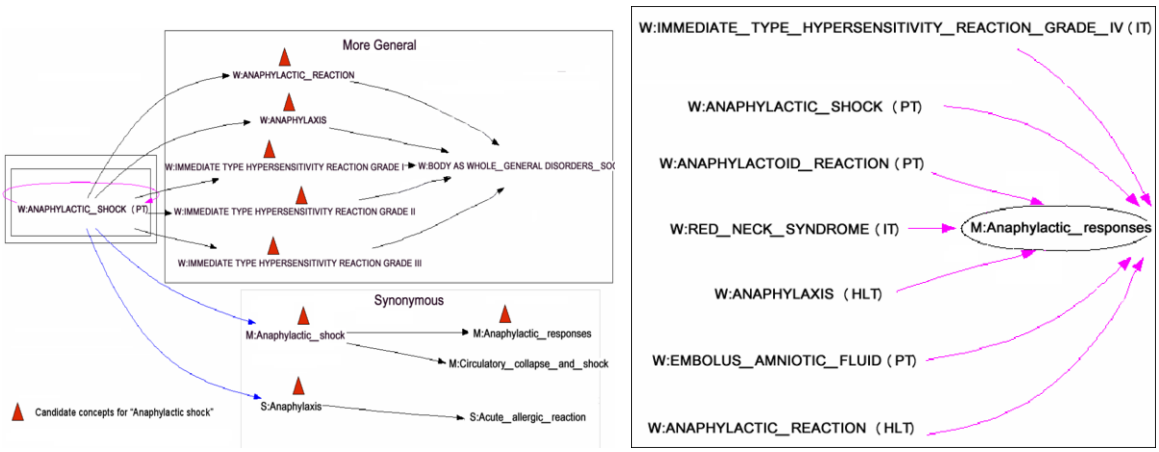
Method 1 uses the Pubmed web service without any refinement, just as the pharmacovigilant will use PubMed. Method 2 uses our own web service VigiPubmed that builds a normalized query for pharmacovigilance. The MedDRA terms are first normalized with UMLS and matched on the MeSH thesaurus [11], then the Pubmed query is built similarly to the one previously proposed [5]. Both methods 3 and 4 also use the VigiPubmed web service. For method 3, the queries are developed successively with all the terms of the MeSH subhierarchy of the adverse effect entry. For method 4, PharmARTS.is first called in order to extend the initial MedDRA entry terms with the OntoEIM ontology, then the group of MedDRA terms is processed with VigiPubmed. The evaluation system is a simple web form application developed in php that integrates the different web services written on a java/axis2 platform.

The relevance of the selected articles is evaluated independently by two pharmacovigilance experts; then the results are compared and in case of discordance, discussed until consensus. A relevant publication is a publication that contributes to the analysis of the ADR case, that helps its comprehension; it is likely to be cited in the literal case report written by a pharmacovigilant.

In method 4, the OntoEIM extension consists in searching OntoEIM for concepts similar to the label used by the pharmacovigilant. We begin by identifying one or several terms associated to the label (WHO-ART term or the SnomedCT and MedDRA Synonymous), using PharmARTS. In a second step, we select candidate concepts in OntoEIM designed by these terms as well as the ascendant when the concept itself doesn't allow any grouping (a leaf concept in the ontology). Finally, the method returns the union of concepts subsumed by all candidate concepts. For example, for the label "Anaphylactic shock" we identify three terms: 1) Anaphylactic shock from Who-ART, 2) Anaphylactic_shock from MedDRA and 3) Anaphylaxis from SnomedCT, these terms are associated with concepts in the OntoEIM ontology and we can therefore carry out three queries with PharmARTS returning three sets of concepts designed by terms from the various terminologies. Figure 2 shows a graphical representation of the organization of these concepts within the ontology.

Example : Anaphylactic shock

List of grouped terms : (ANAPHYLACTIC_REACTION, ANAPHYLACTIC_SHOCK, ANAPHYLACTOID_REACTION, ANAPHYLAXIS, EMBOLUS_AMNIOTIC_FLUID, IMMEDIATE_TYPE_HYPERSENSITIVITY_REACTION_GRADE_I, IMMEDIATE_TYPE_HYPERSENSITIVITY_REACTION_GRADE_II, IMMEDIATE_TYPE_HYPERSENSITIVITY_REACTION_GRADE_III, IMMEDIATE_TYPE_HYPERSENSITIVITY_REACTION_GRADE_IV, RED_NECK_SYNDROME)



(a) *(b)*

Figure 2-Graphical illustration of the ontological extension. (prefix W for Who-ART, M for MedDRA and S for SnomedCT) a) Candidate concepts for term "Anaphylactic shock" b) List of subsumed concepts for candidate concept "Anaphylactic_responses"

Results

The study was performed on the following entry points, based on cases or questions transmitted to the pharmacovigilance department. There are 13 cases grouped in 8 groups which correspond to 16 queries on Pubmed : 1) ADR=lupus, Drugs=statin (atorvastatin, fluvastatin pravastatin rosuvastatin and simvastatin), 2) ADR=purpura, hematoma, petechia or thrombocytopenia, Drug=levetiracetam 3) ADR=anaphylaxy , Drug=rituximab 4) ADR=neutropenia, Drug=ciprofloxacin 5) ADR=hair disorder, Drug=valproate or clonazepam, 6) ADR=Lyell Syndrome, Drug=paracetamol 7) ADR=thrombocytopenia, Drug=infliximab, 8) ADR=Aseptic meningitis, drug=sulfamethoxazole trimethoprim. For instance, in the first group, one can see that there are 5 cases corresponding to the 5 drugs. In the second group, one can see that there is only one case but corresponding to 4 different queries (4 possible terms for the ADR).

The choice of these queries was deliberate in order to check various situations : drugs largely prescribed and with extensive publications (like statins), drugs not widely prescribed (levetiracetam), ADRs that are also indications for the drug (neutropenia and ciprofloxacin), ADRs not clearly defined (purpura, hematoma, petechia or thrombocytopenia), and an expected ADR (aseptic meningitis and sulfamethoxazole trimethoprim). Table 1 displays quantitative results of the comparative study based on 13 reported ADR-drug cases. In the domain of bibliographic review the gold standard is not easily defined, in our study we use a relative reference to calculate recall and precision. For each given case, the reference is the set of relevant abstracts selected by any of the four methods.

Main publication types are "case reports" for 501, unspecified "journal article" for 421, "clinical trial" for 378, "review" for 214, "comparative study" for 96, "letter" for 75, etc.

Table 1- evaluation results of the four search methods

Method	Pubmed	VigiPubmed		
	Method 1	Method 2 Alone	Method 3 + Mesh extension	Method 4 + ONTOEIM extension
Abstracts	480	184	218	1780
Relevant Abstracts	60	48	64	74
Recall	63%	51%	67%	78%
Precision	13%	26%	29%	4%

Compared to the direct request on PubMed, VigiPubmed works as a filter as the 184 VigiPubmed abstracts are is a subset of the 480 Pubmed abstracts. Hence VigiPubmed does improve the precision however other tools are needed to get a better recall. The best recall is provided by "VigiPubmed + OntoEIM", the ontology-based method (78%). In four cases out of 13, this method returns relevant publications when the others methods return no results. For instance, in the case of anaphylactic shock with rituximab, methods 1 to 3 retrieve 8 abstracts but none of these abstracts is relevant while, for method 4, 34 articles are retrieved, and 6 of them, associated with the "RED_NECK_SYNDROME" term, normalized in "hypersensitivity", are selected by the reviewers to be relevant.

Extending the Pubmed request with the ontology causes a loss of precision (4%), this result is expected when we enlarge the context of the bibliographic search, however, for this study, our focus is on the recall in order to demonstrate the "plus value" on the ontological approach in terms of the retrieval of relevant paper. In the future, we can imagine various methods

to browse the results in an efficient manner for increasing the precision: scoring and filtering based upon Mesh heading and qualifier, graphical interface associated to the ontology, etc.

Discussion and conclusion

As of today, we don't have a good evidence for what constitutes an effective search strategy for adverse effects in the medical literature. As the combination of an ADR and a drug is infinite, any method that could normalise and help this search is useful. Moreover, interrogation should not be expert-dependant and should be reproducible whoever realizes it. In the current practice, pharmacovigilance experts start by choosing a term for the adverse effect studied. When the effect is correctly described and well-known, as the choice is easy, the results are easy to analyze, but when few or no documentations are found in the searched databases, experts extend the term to a close adverse effect. The automatic extension to these close terms may therefore be very useful.

Considering the results of this preliminary study, the best way to achieve the bibliographic review on PubMed in terms of cost (that is the number of abstracts to be read) and efficiency (that is the number of relevant articles) is a combination of the different methods. In a first step, the request can be done with the VigiPubmed web service directly or with the MESH extension, the result of the program is a list of articles reduced of about 60 % with regard to a direct request on PubMed whereas recall are 51%, 67% versus 63%. In a second step, when the first request returns no result and also to enlarge and complete the bibliographic review, the request is extended with terms provided by the ontology. The contribution of the ontology is to define a concept into a context by relationships to the others. For instance, a question about thrombocytopenia related to infliximab returns two publications but three additional publications, retrieved with the ontology-based method, reporting another close haematological effect, were useful for the analysis and the comprehension of the case studied.

Another major issue is when the potential adverse effect studied is also a condition or a disease that the drug is used to treat. The best example is neutropenia occurring after antibiotics. The addition of a term such as adverse effect, drug toxicity is not always useful as authors do not always add a specific term for pharmacovigilance data. Global recommendations about pharmacovigilance publications (Drug safety public ERICE) should be applied more systematically [13]. In this first study we adopt the point of view of ADRs. In a future second phase, a similar approach could be applied with the point of view of drugs and active ingredients.

Acknowledgments

The VigiTermes project is funded by the French Agence Nationale pour la Recherche (ANR) ANR-07-TECSAN-026.

References

[1] Bright RA, Nelson RC. Automated support for pharmacovigilance: a proposed system. Pharmacoepidemiol Drug Saf. 2002;11(2):121-5.

[2] Amardeilh F, Bousquet C, Guillemin-Lanne S, Wiss-Thebault M, Guillot L, Delamarre D, Lillo-Le Louet A, Burgun A. A knowledge management platform for documentation of case reports in pharmacovigilance. Stud Health Technol Inform. 2009;150:517-21

[3] Alecu I, Bousquet C, Jaulent MC (2008). A case report: using SNOMED CT for grouping Adverse Drug Reactions Terms. BMC Med Inform Decis Mak. 2008 Oct 27;8 Suppl 1:S4.

[4] Jaulent MC, Alecu I. Evaluation of an ontological resource for pharmacovigilance. In: medical infomatics in a united and healthy europe. K.-P. Adlassning et al. (Eds); IOS Press 2009; pp522-6

[5] Garcelon N, Mougin F, Bousquet C, Burgun A. Evidence in pharmacovigilance: extracting Adverse Drug Reactions articles from MEDLINE to link them to case databases. Stud Health Technol Inform 2006;124:528-533

[6] Loke YK, Price D, Herxheimer A; Cochrane Adverse Effects Methods Group. Systematic reviews of adverse effects: framework for a structured approach BMC Med Res Methodol. 2007;7:32.

[7] Golder S, Loke Y, McIntosh HM. Poor reporting and inadequate searches were apparent in systematic reviews of adverse effects. J Clin Epidemiol. 2008;61(5):440-8.

[8] Brown EG. (2002) Effects of coding dictionary on signal generation: a consideration of use of MedDRA compared with WHO-ART. *Drug Saf* 2002; 25(6):445-52.

[9] Bousquetagier G, LLe Louët C, Le Beller C,Venot A, Jaulent M.C, (2005) Appraisal of the MedDRA conceptual structure for describing and grouping adverse drug reaction. Drug Saf 2005; 28(1):19-34.

[10] Alecu I, Bousquet C, Degoulet P, Jaulent MC (2007). PharmARTS: Terminology Web Services for Drug Safety Data Coding and Retrieval. In : Medinfo 2007. Ed : Kuhn KA, Warren JR, Leong TY. Stud Health Technol Inform. 2007; 129:699-704.

[11] Bodenreider O, Nelson SJ, Hole WT, Chang HF. Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies. Proceedings / AMIA, Annual Symposium. 1998:815–9

[12] Kelly WN, Arellano FM, Barnes J, Bergman U, Edwards RI, Fernandez AM, Freedman SB, Goldsmith DI, Huang KA, Jones JK, McLeay R, Moore N, Stather RH, Trenque T, Troutman WG, van Puijenbroek E, Williams F, Wise RP; International Society for Pharmacoepidemiology; International Society of Pharmacovigilance. Guidelines for submitting adverse event reports for publication. Drug Saf. 2007;30(5):367-73

Address for correspondence

Denis Delamarre, DIM - CHU de RENNES, 2 rue H. Le Guillou 35000 Rennes, France, Email : Denis.Delamarre@chu-rennes.fr