PHARMACOEPIDEMIOLOGY AND DRUG SAFETY 2013; **22**: 1189–1194
Published online 12 August 2013 in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/pds.3493

ORIGINAL REPORT

Automatic detection of adverse events to predict drug label changes using text and data mining techniques†

Harsha Gurulingappa[1‡], Luca Toldo[2*,‡], Abdul Mateen Rajput[2,3], Jan A. Kors[4], Adel Taweel[5] and Yorki Tayrouz[2]

[1]*Molecular Connections Pvt. Ltd., Bangalore, India*
[2]*Merck KGaA, Darmstadt, Germany*
[3]*Bonn-Aachen International Center for Information Technology [B-IT], University of Bonn, Germany*
[4]*Erasmus University Medical Center, Rotterdam, Netherlands*
[5]*School of Medicine King's College, London, UK*

ABSTRACT

Purpose The aim of this study was to assess the impact of automatically detected adverse event signals from text and open-source data on the prediction of drug label changes.
Methods Open-source adverse effect data were collected from FAERS, Yellow Cards and SIDER databases. A shallow linguistic relation extraction system (JSRE) was applied for extraction of adverse effects from MEDLINE case reports. Statistical approach was applied on the extracted datasets for signal detection and subsequent prediction of label changes issued for 29 drugs by the UK Regulatory Authority in 2009.
Results 76% of drug label changes were automatically predicted. Out of these, 6% of drug label changes were detected only by text mining. JSRE enabled precise identification of four adverse drug events from MEDLINE that were undetectable otherwise.
Conclusions Changes in drug labels can be predicted automatically using data and text mining techniques. Text mining technology is mature and well-placed to support the pharmacovigilance tasks. Copyright © 2013 John Wiley & Sons, Ltd.

KEY WORDS—adverse event; text mining; machine learning; signal detection; pharmacoepidemiology

Received 9 January 2013; Revised 28 June 2013; Accepted 4 July 2013

INTRODUCTION

Worsening of a drug's risk-to-benefit ratio can lead regulatory authorities to take actions aiming at mitigating or withdrawing the use of the drug. A very first step in such actions is the issue of a safety warning that has modest impact[1] on the usage of a drug. One important way by which regulatory authorities may act on safety warning issuance is through changes in drug labels. Drug labels (also referred as leaflets) contain information provided by manufacturers on the drug usage, characteristics and safety.[§] Therefore, changes in safety labels can have a strong impact on the way drugs can be used by physicians and patients.

Adverse drug events are the core data that drive the regulator's decision-making process. While clinical studies and spontaneous adverse reporting systems provide the major source of adverse drug events, scientific literature provides a secondary but a valuable data source. The detection of potential adverse drug events from literature is often a manual process done by pharmacovigilance experts mainly using literature databases and information retrieval systems.

Several examples of successful application of text mining techniques for adverse event detection are available.[2–4] A text mining system originally developed to detect semantic relations between nouns occurring in English sentences[2] was shown to perform well in detecting drug–drug interactions in biomedical text[3] upon customization. The same text mining

*Correspondence to: L. Toldo, Merck KGaA, Frankfurterstr 250. Email: luca.toldo@merckgroup.com
†PRIOR POSTINGS AND PRESENTATIONS: This work has never been published or presented before, neither in Journals nor conferences or posters.
‡Harsha Gurulingappa and Luca Toldo have equally contributed as first authors.
§http://www.mhra.gov.uk/Howweregulate/Medicines/Labelspatientinformationleafletsandpackaging/index.htm.

system[3] was recently applied to detect potential adverse drug events with good results.[4] To the best of our knowledge, neither a benchmark of different text mining technologies for adverse drug event detection nor an evaluation of the practical impact of automatically extracted information on pharmacovigilance processes have been reported so far.

Thus, the aim of this paper is to conduct a study on extraction of adverse event signals from open-source adverse event reports and MEDLINE case reports, and subsequent detection of drug safety signals using statistical approaches. Predicted signals were used to assess their potential impact on the prediction of changes in drug safety labels.

METHODS

Extraction of adverse events from MEDLINE

Characteristics of corpora. The ADE-EXT corpus[5] was used as a reference for training and evaluation of performances of text mining system applied for adverse event extraction from MEDLINE case reports. The corpus contains 2972 abstracts randomly sampled from the MEDLINE query that returned nearly 29 800 abstracts.

> *"adverse effects"[sh] AND (hasabstract[text] AND Case Reports[ptyp]) AND "drug therapy"[sh] AND English [lang] AND (Case Reports[ptyp] AND ("1"[PDAT] : "2009/12/31"[PDAT])).*

Sampled abstracts were annotated with drugs and conditions as well as relationships occurring between them at the sentence level.[6] Annotations were performed by three annotators followed by systematic harmonization to remove conflicting and error-prone annotations. The corpus contains 7528 drugs and 9517 condition annotations. There are 6701 *positive* and 5345 *negative* relations between drugs and conditions occurring within the sentences.

The ADE-EXT corpus contains annotations of "drug" related to "condition" as *positive* if and only if the drug is causing/worsening the condition, e.g.sentence-1 (S1): Tamoxifene was used to treat breast cancer.sentence-2 (S2): Tamoxifene caused liver failure in a patient.In ADE-EXT corpus, the *positive* annotation would be "S2: Tamoxifene - liver failure" and the negative annotation would be "S1: Tamoxifene - breast cancer".

Apart from the ADE-EXT corpus, a set of 1 417 079 articles was extracted from MEDLINE that were published before 2009/01/01. This corpus, namely MEDLINE-2008, was used for large-scale extraction of adverse events from text that can be subjected to signal detection.

Text mining systems applied for adverse event detection. **Java Simple Relationship Extraction (JSRE)** is an open-source publicly available[¶] machine learning-based relationship extraction system.[2] JSRE was trained over ADE-EXT corpus for the detection of relations between drugs and adverse events co-occurring within sentences of case reports. JSRE requires pretagging of drugs and adverse events in sentences prior to relation detection.

Peregrine is a publicly available[‖] dictionary-based named entity recognition system.[7] Peregrine was loaded with chemical and medical dictionaries, i.e. DrugBank[**] and MedDRA[††], respectively, for the identification of sentence-level co-occurring drug and condition named entities.

Training the text mining systems and large-scale adverse event extraction. From the previous experiments, it has been demonstrated that training the JSRE on the ADE-EXT corpus and evaluation can deliver relationship extraction results with F-score of 0.87 (Precision = 0.86 and Recall = 0.89). Gurulingappa *et al.,* 2012[5] give a brief insight on methodological aspects and performance of the relationship extraction. Since the JSRE showed high confidence in accurate adverse event (*positive* labeled) relationship extraction from text, it was used for large-scale relation extraction from MEDLINE.

In the first step, Peregrine was applied over the MEDLINE-2008 corpus for tagging drugs and condition names co-occurring within the sentences. A JSRE model trained over the ADE-EXT corpus was applied for classification of relationships between co-occurring drugs and conditions tagged within sentences of the MEDLINE-2008 corpus. As a result of relationship classification, 165 680 adverse event relationships were extracted between 1611 drugs and 5079 conditions from 70 024 articles.

Extraction of adverse events from open-source data

Medicines and Healthcare products Regulatory Agency (MHRA) label changes dataset. The MHRA dataset consists of the 62 drug label changes (for 29 drugs)[‡‡] that have been reported from March to December 2009 by the UK pharmacovigilance authorities

[¶] http://hlt.fbk.eu/en/technology/jSRE.
[‖] https://trac.nbic.nl/data-mining/.
[**] http://www.drugbank.ca/.
[††] http://www.meddramsso.com/.
[‡‡] MHRA personal communication.

DOI: 10.1002/pds

in Drug Safety Updates (DSU).[§§] It is very important to understand that only a proportion of product label changes are communicated via DSU, and as a result this cannot be used as a complete list of changes that have been made. Communication via DSU is based on major changes to the benefit/risk balance for a medicine; this includes any new major safety information, major updates to prescribing advice and even reminders. Therefore, many minor amendments will not be included in DSU. The MHRA may communicate some of the other minor product label changes through other means (drug alerts and variation UKPARs); these again may only constitute a proportion of all changes made to a product label. These 62 drug label changes (62 drug-adverse event pairs) served as a standard reference for comparing the results of machine prediction.

Yellow Cards dataset. The Yellow Card Scheme is run by the MHRA in the United Kingdom, established in 1964. This is the main reporting system for spontaneous "suspected" Adverse Drug Reactions (ADR) of the United Kingdom. Custom generated MS Excel files of the publicly available Drug Analysis Prints[¶¶] were obtained for 29 MHRA label-changed drugs from the Yellow Cards database dated between 01/07/1963 and 31/12/2008. Each Excel file contains cumulative counts of outcomes for each drug-adverse event coded using MedDRA Preferred Term. It is important to note that causality is not proven for these ADR reports – reporters are encouraged to report spontaneous "suspected" ADRs, but the reporter does not have to be sure that the drug caused the reaction – a mere suspicion will suffice. The limitations of a spontaneous reporting scheme such as the Yellow Card Scheme include an unknown level of underreporting. ADR reporting rates may be influenced by seriousness of reaction, their ease of recognition or publicity about a drug. Yellow Card data cannot be used to determine incidence of a particular ADR because denominator data are not available. We are basing all the product information updates on DSU alone. The monthly bulletin DSU is used to communicate some of the issues MHRA has identified; however, note that not every single product information update is communicated this way. Publications communicated via DSU articles can also be identified from other sources of evidence, such as epidemiological reviews, literature or other studies, so it is important to note that Yellow Card data is not the only source. We therefore do not ascertain that there were only 63 product label changes in 2008, but these were the ones that we sought to automatically identify.

FAERS dataset. The FDA Adverse Event Reporting System (FAERS)[‖‖] is a database that contains information about adverse drug events submitted to FDA. The data of the FAERS database is freely downloadable in different formats and contains data from January 2004, updated quarterly thereafter. Each file contains all the individual reports registered by FAERS during the specified period for any drug. The complete dataset was downloaded in SGML format. For each of the 29 MHRA label-changed drugs, we extracted the cumulative counts of outcomes for each reaction submitted before 2009.

SIDER dataset. SIDER is a publicly available*** computer-readable side effect resource that connects 888 drugs to 1450 side-effect terms. It has been constructed manually from the summary of product leaflets of each drug. SIDER version 1.0 was used in the study for signal detection.[8] Although the release date is 6 November 2009, the data present in SIDER has been generated before 2009.[†††] For each of the 29 MHRA label-changed drugs, adverse events were extracted from the SIDER product leaflets. SIDER provides drug-adverse event pairs in the form of table with frequency information extracted from product leaflets.

MEDLINE dataset. This dataset contains drug-adverse event pairs extracted by text mining from the MEDLINE-2008 corpus (see Section 2.1). Only the 29 MHRA label-changed drugs were used.

Prediction of changes in drug safety labels

Three datasets used in the current study, i.e. FAERS, Yellow Cards and MEDLINE for 29 drugs of interest, were subjected to the signal detection process. Drug safety signals were detected using the Multi-item Gamma Poisson Shrinkage (MGPS) method[9] extended to the multiple comparison framework[10] and implemented within the publicly available R[‡‡‡] PhViD package version 1.0.3.[11] MGPS is one of the best statistical methods

[§§]http://www.mhra.gov.uk/Safetyinformation/DrugSafetyUpdate/DrugSafetyUpdatePDFarchive/index.htm.
[¶¶]http://yellowcard.mhra.gov.uk/.
[‖‖]http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/.
[***]ftp://sideeffects.embl.de/SIDER/2009-11-06/.
[†††]information found in the label_mapping.tsv file of SIDER.
[‡‡‡]http://cran.r-project.org/.

DOI: 10.1002/pds

used in pharmacovigilance for disproportionality analysis (to estimate disproportionality in the frequency of adverse events) and allows detecting signals. Unlike other datasets, SIDER contains information extracted from product leaflets and does not require statistical signal generation. Therefore, SIDER was not subjected to MGPS-based signal generation and all the drug-adverse event pairs extracted from SIDER were treated as signals. Signals detected from different datasets were manually compared against 2009 MHRA drug label changes. Figure 1 depicts the methodological workflow applied for drug-adverse event extraction, signal detection and the prediction of drug label changes.

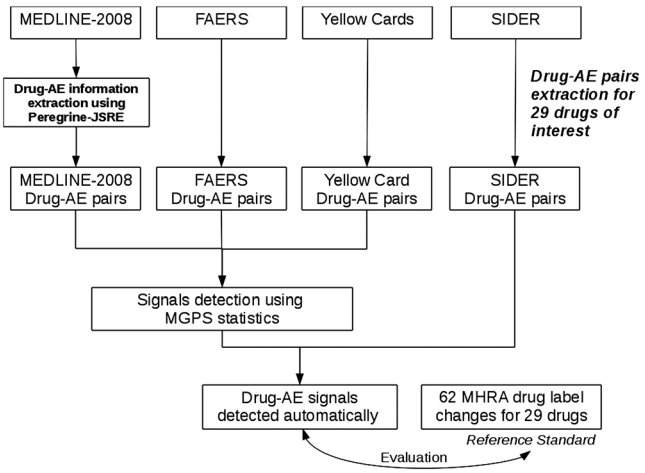

Figure 1. Workflow applied for drug-adverse event extraction from various sources, signal detection and prediction of drug label changes

RESULTS

Prediction of label changes

Upon application of MGPS over four datasets, drug-adverse event pairs that were qualified as valid drug safety signals were extracted and manually compared to 62 instances in the MHRA label change dataset. Among 62 instances, 47 instances (i.e. 76%) could be predicted through any one of the datasets (*see Appendix 1*). This indicates that 76% of the drug label changes could be automatically predicted through signals generated from the open-source data. The reason for 24% unpredicted label changes was due to the lack of data in the repositories used (i.e. MEDLINE, AERS, Yellow Cards or SIDER) due to which no signals could be predicted for the missing drugs.

Table 1 provides the counts of extracted adverse events and signals generated for 29 drugs from

Table 1. Number of adverse events extracted from different resources for 29 drugs under study. Numbers within brackets indicate the number of signals (i.e. after MGPS-based filtering) generated for each drug. For SIDER, number of extracted adverse events equals number of signals since no MGPS-based filtering was applied

Drug	FAERS	Yellow Cards	SIDER	MEDLINE-2008
Alendronic acid	150 (26)	677 (105)	41 (41)	No data
Bromocriptine	No data	337 (54)	57 (57)	96 (16)
Cabergoline	695 (177)	336 (44)	69 (69)	22 (5)
Carbamazepine	1873 (388)	1103 (179)	11 (11)	219 (11)
Divalproex sodium	319 (59)	No data	No data	10 (4)
Efalizumab	173 (42)	73 (11)	No data	20 (5)
Evra	No data	No data	No data	No data
Exenatide	192 (87)	No data	No data	1 (1)
Felbamate	69 (34)	No data	No data	9 (3)
Fentanyl	1423 (263)	467 (91)	220 (220)	35 (13)
Gabapentin	1562 (247)	706 (80)	294 (80)	64 (8)
Ketoconazole	380 (90)	273 (48)	38 (38)	59 (12)
Lamotrigine	1276 (227)	915 (106)	206 (206)	64 (10)
Modafinil	384 (91)	146 (16)	118 (118)	10 (2)
Natalizumab	1498 (321)	42 (4)	No data	6 (2)
Nicorandil	70 (15)	427 (53)	No data	No data
Nitrous oxide	112 (19)	95 (11)	No data	No data
Oxcarbazepine	522 (70)	167 (11)	158 (158)	27 (5)
Pamidronic acid	39 (4)	310 (50)	113 (113)	No data
Pegylated interferon	27 (3)	No data	No data	No data
Pergolide	13 (1)	237 (49)	249 (249)	20 (9)
Pregabalin	458 (88)	651 (84)	453 (453)	14 (3)
Rimonabant	37 (5)	438 (74)	No data	4 (1)
Rituximab	1227 (312)	360 (82)	No data	131 (24)
Telbivudine	39 (15)	No data	No data	No data
Tiagabine	1 (1)	164 (20)	196 (196)	10 (1)
Varenicline	329 (58)	817 (123)	167 (167)	9 (7)
Zoledronic acid	1064 (243)	385 (64)	No data	No data
Zonisamide	635 (146)	109 (10)	No data	23 (5)

DOI: 10.1002/pds

different resources. It was striking to observe that (i) MGPS-based signal detection helped in substantial filtration of drug-adverse event pairs thereby reducing a lot of manual work and (ii) MGPS-based filter did not discard any valuable information (drug-adverse event pairs) that were reported in the 2009 label change standard reference. Signals generated through text/data mining that were not reported as 2009 label changes (i.e. not present in the reference standard) were not termed as "false positives" but rather treated as signals requiring manual validation. The reasons are:

a) Many predicted signals that did not lead to label changes in 2009 were adverse events already documented in the drug labels.
b) For the remaining predicted signals, there was no evidence of manual evaluation available to render them as "*false*".

Table 2 provides the counts of label changes that could be predicted using each dataset.

Figure 2 depicts overlaps in label changes that could be predicted using different datasets.

As reported in Table 3, it was striking to observe that by using a text mining application we were able to identify four drug adverse event signals (label changes) that were not detected by the other sources. This shows the potential of text mining applications for ability to detect novel signals from text and their subsequent support to the pharmacovigilance activities.

Table 2. Counts of drug label changes that could be identified using each dataset

Dataset	Total no. of label changes predicted	Unique label changes predicted by the current dataset alone
FAERS	30	5
MEDLINE	19	4
SIDER	16	2
Yellow Card	29	3
Combined	**47** (*after removing duplicates*)	-

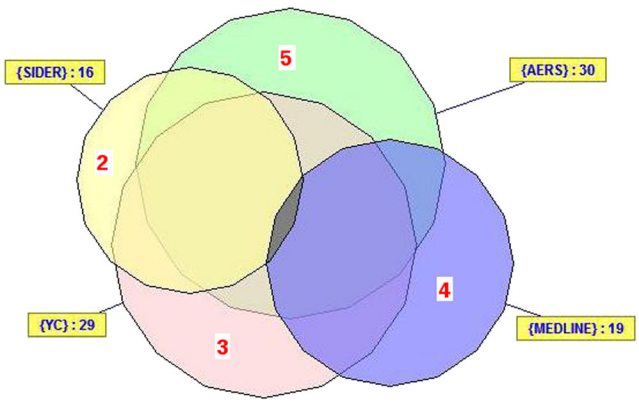

Figure 2. Overlaps in label changes predicted using different datasets. Numbers in *red* color indicate unique label changes predicted from each resource

Table 3. Drug-adverse event signals detected from MEDLINE case reports and not detected from other datasets

Drug	Adverse event
Modafinil	Suicidal ideation
Zonisamide	Suicidal ideation
Felbamate	Suicidal ideation
Felbamate	Suicidal behaviour

CONCLUSIONS

This work demonstrates that up to 76% of drug label changes could be predicted through data mining methods using publicly available data. Furthermore, text-mining solutions can deliver promising results by detecting unreported adverse events and can add value to the safety signal instantiation processes. The Peregrine-JSRE hybrid system that delivered the highest precision and recall on the ADE-EXT benchmark corpus was able to detect uniquely four adverse drug events that were otherwise not found in the other databases. This demonstrates the usability of well-tuned text-mining software to support an important pharmacovigilance task.

The results show a promising initial outcome of using off-the-shelf text-mining tools to demonstrate their value for the drug adverse-event detection. It is useful to benchmark the capabilities of different text mining systems against a common corpus and determine their abilities to detect potential safety signals. Nevertheless, there are few limitations that authors wish to address in the future studies. For instance, information extraction has been limited to MEDLINE case reports. The model exhibited here may not adapt straightforward to full-text articles or non-case reports. Therefore, performance evaluation and training requirements of entity recognition and relation extraction on a broad spectrum of corpora need to be addressed. Further development of dedicated text mining models, training the tools on different corpora and/or fine-tuning of system parameters will be systematically investigated. Furthermore, the relative abundance of inter-sentence relations vs. intra-sentence relations should be assessed, and methods be selected that detect inter-sentence relations. In the current study, prediction of MHRA drug label changes has been

DOI: 10.1002/pds

addressed. Similarity, ability to predict drug label changes for various other regulatory authorities needs to be investigated. Therefore, application of data mining and text mining technologies could revolutionize the way in which decisions on drug safety could be taken by safety experts in the near future.

ETHICS

None of the datasets had content that could be used to identify a specific individual; therefore, this study does not raise any ethical concern.

CONFLICT OF INTEREST

LT, YT and AMR have been sponsored by Merck KGaA, a chemical and pharmaceutical company.

KEY POINTS

- Approximately 76% of the drug label changes in this study were predicted using the public data.
- A shallow linguistic relation extraction method enabled precise detection of adverse drug events in abstracts of MEDLINE case reports.
- 6% of the drug label changes could be predicted only through text mining and not through the other resources used in the study indicating the value of text mining to contribute to computational pharmacovigilance.

ACKNOWLEDGEMENTS

This work would have not been possible without the outputs from the Yellow Card Scheme, provided by the UK Medicines and Healthcare products Regulatory Agency.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix 1. Detailed result of the MHRA drug label change benchmarking

REFERENCES

1. Piening S, Haaijer-Ruskamp FM, de Vries JT, *et al.* Impact of Safety-Related Regulatory Action on Clinical Practice: A Systematic Review. *Drug Saf* 2012; **35**(5): 373–385.
2. Giuliano C, Lavelli A, Pighin A, *et al.* FBK-IRST: kernel methods for semantic relation extraction, In *Proceedings of the 4th International Workshop on Semantic Evaluations.* Association for Computational Linguistics: Prague, Czech Republic, 2007.
3. Chowdhury FM, Lavelli A. Drug-drug interaction extraction using composite kernels, In 1st Challenge task on Drug-Drug Interaction Extraction (DDIExtraction 2011). Huelva: Spain, 2011.
4. Gurulingappa H, Rajput A-M, Toldo L. Extraction of Adverse Drug Effects from Medical Case Reports. in *International Conference Biomedical Ontologies*, M. Courtot, Editor. Graz, Austria, 2012.
5. Gurulingappa H, Mateen-Rajput A, Toldo L. Extraction of Adverse Drug Effects from Medical Case Reports. *J Biomed Semantics* 2012; **3**: 15.
6. Gurulingappa H, Rajput AM, Roberts A, *et al.,* Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. *J Biomed Inform* 2012; **45**(5): 885–892.
7. Schuemie M, Jelier R, Kors J. Peregrine: lightweight gene name normalization by dictionary lookup. in *Proceedings of the Second BioCreative Challenge Evaluation Workshop: 23-25 April* 2007. Madrid.
8. Kuhn M, Campillos M, Letunic I, *et al.* A side effect resource to capture phenotypic effects of drugs. *Mol Syst Biol* 2010. 6.
9. DuMouchel W, Bayesian Data Mining in Large Frequency Tables, with an Application to the FDA Spontaneous Reporting System. *Am Stat* 1999; **53**(3): 177–190.
10. Ahmed I, Thiessard F, Miremont-Salamé G, *et al.* Pharmacovigilance Data Mining With Methods Based on False Discovery Rates: A Comparative Simulation Study. *Clin Pharmacol Ther* 2010; **88**(4): 492–498.
11. Ahmed I, Poncet A. PhViD: a R package for PharmacoVigilance signal Detection. 2011.