Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220499

An Adverse Drug Reaction Database for Clinical Use – Potential of and Difficulties with the Summary of Product Characteristics

Birgit EIERMANN[a,b,1] , Daniel RODRIGUES[c], Paul COHEN[c] and Lars L GUSTAFSSON[a,c]
[a] *Department of Laboratory Medicine, Division of Clinical Pharmacology, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden*
[b] *Inera AB, Swedish Association of Local Authorities and Regions, Stockholm, Sweden*
[c] *Swedish Institute for Drug Informatics (SIDI), c/o Division of Clinical Pharmacology, Karolinska Institutet, Karolinska University Hospital, Stockholm, Sweden*

Abstract. Adverse drug reactions (ADRs) for all drugs in Europe are described in the legally approved Summary of Product Characteristics (SmPC). An overview of all ADRs of the patients' drug list can support healthcare staff to link patient symptoms to possible ADRs. We review the possibilities and challenges to extract ADR information from SmPCs and present the development of our semi-automated procedure for extraction of ADRs from the tabulated section of the SmPCs to create a database, named Bikt, which is regularly updated and used at point of care in Sweden. The existence of five major table formats for ADRs used in the SmPCs required the development of different parsing scripts. Manual checks for correctness for all content has to be performed. The quality of extraction was investigated for all SmPCs by measuring precision, recall and F1 scores (i.e. the weighted harmonic mean of precision and recall) and compared with other methods published. We conclude that it is possible to semi-automatically extract ADR information from SmPCs. However, clear technical and content guidelines and standards for ADR tables and terms from drug registration authorities would lead to improved extraction and usability of ADR information at point of care.

Keywords. Adverse drug reaction; ADR-database, clinical decision support; data extraction; knowledge bases; Summary of Product Characteristics

1. Introduction

Adverse drug reactions (ADRs) are defined by the World Health Organization (WHO) as 'a response to a medicine which is noxious and unintended, and which occurs at doses normally used in man'. ADRs are costly and they are especially common when treating patients with multiple diseases prescribed a high number of drugs. About 80% are dose-dependent and can be avoided [1,2].

[1] Corresponding Author, Birgit Eiermann, Box 17703, 118 93 Stockholm, Sweden; E-mail: birgit.eiermann@inera.se.

In Europe, ADRs for approved drugs are described in section 4.8. of the Summary of Product Characteristics (SmPC) which are approved by the national medical product agencies (MPA) or European Medicines Agency (EMA). EMA publishes guidelines for the SmPC structure for section 4.8 which should contain a tabulated section of ADRs, listed according to their frequency and using the system for organ classification (SOC), where they occur, as defined by MedDRA (Medical Dictionary for Regulatory Activities). However, provided templates for the SmPC offer little guidance for the layout of the table, no support for selection of a controlled vocabulary for ADRs and provide no other guidelines for the description of ADRs. This causes a wide variation in table formats, inconsistent use of terms and inconsistent linking of ADR terms to specific SOC terms with challenges both for machine and human reading [3].

Integrated access to information on ADRs at point-of-care, could assist healthcare staff to choose and dose drugs appropriately for an individual patient. A summary of all the potential ADRs of all patients´ drugs could help physicians by providing a good overview of ADRs to be considered for a patient. An ADR database integrated into an electronic health record system (EHR) could serve as a clinical decision support for healthcare personnel.

This paper outlines the development of processes and methods to build the ADR database "Bikt" in Sweden. It describes various structures of existing SmPC ADR tables, the extraction and updating process of ADR information as well as the procedure for distribution of the database to EHR-systems to be used at point of care. It also points out which table structure is most suitable for automated extraction of ADRs from SmPC documents available in Europe after having performed an experiment. [4]

2. Development of the ADR database and its distribution to EHR systems

The Bikt database contains ADR information for over 95% of all approved drugs on the Swedish market (a total of 5429 SmPCs). It is updated regularly on a monthly basis, distributed centrally, and integrated into several major EHR.

The development process of the ADR database Bikt comprises of the following steps:

- Data import: retrieval of SmPC documents from various websites and their linkage to national drug entities.
- Extraction process: Step 1: transformation of SmPC documents into pure text format with parts of the layout information retained and split into separate text segments based on the chapter and section structure of the SmPC. Step 2: extraction of ADRs using 5 principle parsing scripts. Each SmPC is parsed with each script and the script that gives the best results, is used to classify the ADR table format. The five parsing scripts were developed, in an iterative process using a heuristic method, based on visual inspection of various SmPCs
- Manual editing and approval: a custom-built editing tool is used by pharmacists to control and correct the extracted ADRs following a standardized operation procedure. Additionally, the linkage to the national drug identifier is checked

before approval. Corrections of the extracted material can be e.g removal of terms (e.g. “has been reported”), or footnotes which are not included in Bikt.

- Export and distribution to point of care: Exported ADR tables are integrated into a national database (Sil-database; Sil = Swedish information services for drugs) which is distributed to all counties in Sweden and implemented in all major EHR.
- Update: New SmPCs are added to the database and existing SmPCs are checked for changes in the ADR section.

3. Quality of Data Extraction

In an experiment, we quantified the suitability of the various table formats for correct extraction using the subset of SmPCs which already had been manually checked and approved. We technically compared the corrected subset of SmPCs with a newly extracted version of the same SmPCs. The subset consisted of more than 3200 SmPCs. The performance of our models for extraction of the different table types were evaluated using standard metrics methods: Precision (i.e. true positive/(true positive + false positive)), Recall (i.e. true positive/(true positive + false negative)) and F1-score (i.e. the weighted harmonic mean of precision and recall) [5]. True positive (TP) terms are defined as terms which do exist in the extracted version and in the manually corrected version. False positive (FP) terms are defined as terms which exist in the extracted version but not in the manually corrected version. False negative (FN) terms are defined as terms which do not exist in the extracted version but exist in the manually corrected version. The same applies for the analysis of the combination of SOC and frequency. Two analyses of precision, sensitivity and F1 score were performed to determine best extraction results for the various table formats:

1. Precision, recall and F1 score for “All ADR terms”; which ADR terms are included or missing in the newly extracted version compared to the manually corrected one.
2. Precision, recall and F1 score for “The combination of SOC and frequency”; each SOC is combined with one to many frequencies in an SmPC; this combination is compared in the newly extracted version with the manually corrected one.

During the development of the parsing scripts using an iterative process, we identified 5 table types:

- DFSU - frequencies in the rows followed by indented SOCs and ADRs
- LFSU - frequencies in the column header, and system organ classes followed by ADRs in the rows
- LSFU - SOCs in the column header followed by frequencies and ADRs in the rows
- MSFU - complete tables with SOCs in the rows and frequencies in the column header containing empty table fields where no ADRs exist for certain SOC-frequency combinations
- TSFU - SOCs are displayed per row followed by frequencies and ADRs

(D= Disposition, F = frequency, L = Layout, M = Matrix, S = SOC, T = Table, U = Undesirable effect).
LSFU tables are the most frequent ones used within the SmPC documents (n = 1881).

The results for the extraction of the various table formats are summarized in table 1. The TSFU format (fig. 1) was extracted with the highest precision and recall for both correct ADR terms (0,96 and 0,92, respectively) and the combination of SOC and frequency terms (0,97 and 0,93, respectively).

Table 1. Precision, recall and F1-score calculated for the extraction of adverse drug reaction terms and the combination of system organ class and frequency for all table type variants.

	Extraction of ADR terms			Extraction of combination of SOC and frequency		
Table type (number of SmPCs)	Precision	Recall	F1-score	Precision	Recall	F1-score
DFSU (188)	0,71	0,70	0,70	0,72	0,69	0,7
LFSU (440)	0,85	0,77	0,78	0,41	0,27	0,31
LSFU (1411)	0,66	0,94	0,73	0,92	0,89	0,89
MSFU (549)	0,87	0,58	0,66	0,58	0,43	0,48
TSFU (618)	0,96	0,92	0,93	0,97	0,93	0,94

System organ class	Frequency	Adverse reaction
Immune system disorders	Not known	Hypersensitivity reactions, both local and generalised, including rash*
Psychiatric disorders	Uncommon	Mood swings
Nervous system disorders	Common Uncommon	Headache Dizziness
Vascular disorders	Uncommon	Hot flush

Figure 1. Example of the "table-based" (T) format TFSU: system organ classes (S) are listed in the rows followed by frequencies (F) and undesirable effects (U).

4. Discussion and Conclusions

SmPC documents are important sources of medical information in Europe on approved drugs for healthcare staff. However, these legally approved sources lack standardization of the technical structure as well as of the graphical layout and ADR terminology used. The expressed need of Swedish physicians to have access to an ADR overview resulted in the semi-automated construction of the ADR database, Bikt based on SmPCs of all drugs on the Swedish market and implemented at point of care [6]. Review of the literature showed that there is no other ADR database in Europe being updated regularly and integrated into EHR systems to be used at point of care.

Various table formats led to the development of 5 major parsing scripts. An experiment showed which table format is most suitable for automatic extraction. However, manual

control has to be performed for all SmPCs which is time consuming, costly, prone to introduce mistakes and leads to reduced update frequency. Therefore, the improved use of standards for both table format and terminology within the SmPC guidelines for the ADR tables, and an increased control of the adherence to the standards by the medical product agencies would lead to improvement for both human and machine reading purposes.
Bikt is implemented in over 50% of the counties in Sweden through integration into three major EHR. Future evaluations are planned regarding the use and possible benefits within different groups of healthcare personnel.

Ultimately, information in the SmPCs should be entered according to a defined standard and in a formal and consistently structured way from the beginning to increase their usability as suggested earlier [7]. The development of parsing scripts or other extraction techniques and time-consuming manual control could then be skipped, and information could be used immediately at point of care.

References

[1] Lagnaoui R, Moore N, Fach J, Longy-Boursier M, Bégaud B. Adverse drug reactions in a department of systemic diseases-oriented internal medicine: prevalence, incidence, direct costs and avoidability. Eur J Clin Pharmacol 2000;56:181-6.
[2] Natanaelsson J, Hakkarainen KM, Hägg S, Andersson Sundell K, Petzold M, Rehnberg C, Jönsson AK, Gyllensten H. Direct and indirect costs for adverse events identified in medical records across care levels, and their distribution among payers. Res Social Adm Pharm 2016; pii: S1551-7411:30237-6.
[3] Lamy JB, Ugon A, Berthelot H. Automatic extraction of drug adverse effects from product characteristics (SPC): a text versus table comparison. Stud Health Technol Inform 2016;228: 339-43.
[4] Eiermann B, Rodriguez D, Cohen P, Gustafsson LL. ADR databases for on-site clinical use: potentials of Summary of Products Characteristics. Basic Clin Pharmacol Toxicol. 2021; 28(4):557-567.
[5] Van Rijsbergen C. Information retrieval. Second edition London, UK: Butterworths; 1979.
[6] Rahmner PB, Eiermann B, Korkmaz S, Gustafsson LL, Gruvén M, Maxwell S, Eichler HG, Vég A. Physicians' reported needs of drug information at point of care in Sweden. Br J Clin Pharmacol 2012;73:115-25.
[7] Maxwell S, Eichler HG, Bucsics A, Haefeli WE, Gustafsson LL; e-SPC Consortium. e-SPC - delivering drug information in the 21st century: developing new approaches to deliver drug information to prescribers. Br J Clin Pharmacol 2012;73:12-5.