Journal of Biomedical Informatics xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Decision support environment for medical product safety surveillance

Taxiarchis Botsis [a,*], Christopher Jankosky [a], Deepa Arya [a], Kory Kreimeyer [a], Matthew Foster [a], Abhishek Pandey [a], Wei Wang [b], Guangfan Zhang [b], Richard Forshee [a], Ravi Goud [a], David Menschik [a], Mark Walderhaug [a], Emily Jane Woo [a], John Scott [a]

[a] Office of Biostatistics and Epidemiology, Center for Biologics Evaluation and Research, US Food and Drug Administration, United States
[b] Engility Corporation, United States

ARTICLE INFO

Article history:
Received 4 February 2016
Revised 22 July 2016
Accepted 27 July 2016
Available online xxxx

Keywords:
Natural language processing
Text mining
Network analysis
Post-marketing surveillance
Information retrieval

ABSTRACT

We have developed a Decision Support Environment (DSE) for medical experts at the US Food and Drug Administration (FDA). The DSE contains two integrated systems: The Event-based Text-mining of Health Electronic Records (ETHER) and the Pattern-based and Advanced Network Analyzer for Clinical Evaluation and Assessment (PANACEA). These systems assist medical experts in reviewing reports submitted to the Vaccine Adverse Event Reporting System (VAERS) and the FDA Adverse Event Reporting System (FAERS). In this manuscript, we describe the DSE architecture and key functionalities, and examine its potential contributions to the signal management process by focusing on four use cases: the identification of missing cases from a case series, the identification of duplicate case reports, retrieving cases for a case series analysis, and community detection for signal identification and characterization.

Published by Elsevier Inc.

1. Introduction

Post-marketing passive surveillance is a critical safeguard, providing early evidence for unexpected adverse events (AEs) after the licensure of a medical product. Several systems have been established worldwide to receive, assemble, and analyze AE reports. The U.S. Vaccine Adverse Event Reporting System (VAERS) and the U.S. FDA (Food and Drug Administration) Adverse Event Reporting System (FAERS) are spontaneous reporting systems that collect reports of AEs following immunization and drug exposure, respectively [1,2]. These reports are submitted by patients, healthcare providers, manufacturers, and other interested parties. The reports include narratives of the AEs, which are coded using the Medical Dictionary for Regulatory Activities (MedDRA) terminology. The AE MedDRA codes complement the reported demographic, date, product, and other information.

Once reports are in these systems, statistical algorithms such as the Multi-item Gamma Poisson Shrinker are used to identify drug-AE associations that occur with disproportionate frequency [3]. In addition, physicians review both individual reports and groups of reports to identify and investigate patterns of adverse drug reactions, looking for hallmarks of causal associations such as consistency, temporality, biological plausibility, and dose-response [4]. There are significant limitations to both automated signal detection and medical case review approaches when utilizing spontaneously reported data. Automated signal detection algorithms generally focus on simple bivariate relationships (one drug → one AE), which do not reflect the complexity of individual health events. Although several multivariate signal detection methods have been proposed in the literature, these are not widely used, likely in part due to computational burden [5,6]. Medical case review has its own challenges due to the sheer volume of reports: FAERS receives approximately 770,000 reports and VAERS receives approximately 35,000 reports each year [7]. During the period 2006–2015, the number of reports submitted to VAERS more than doubled compared with the previous ten years, growing from 138,381 to 354,300. Over the same time, FAERS reports almost quadrupled from approximately 2.16 million to 8.03 million.[1] While all serious adverse events are intensively reviewed by Medical Officers (MOs), the volume of reports can make it prohibitive to fully review each individual report, and impossible in some cases to "connect the dots" between disparate reports that reflect a common underlying phenomenon. It is therefore critical to develop tools that will efficiently, effectively, and rigorously assist medical experts and epidemiologists in accomplishing their surveillance duties.

* Corresponding author at: Office of Biostatistics & Epidemiology, Center for Biologics Evaluation and Research, FDA, 10903 New Hampshire Ave, WO71 – 1233, Silver Spring, MD 20993-0002, United States.
E-mail address: Taxiarchis.Botsis@fda.hhs.gov (T. Botsis).

[1] Calculations based on data retrieved from the US VAERS and FAERS systems on 5/31/2016.

http://dx.doi.org/10.1016/j.jbi.2016.07.023
1532-0464/Published by Elsevier Inc.

At the FDA's Center for Biologics Evaluation and Research, we have developed a computational Decision Support Environment (DSE) to facilitate and improve the medical review of spontaneous AE reports. This environment contains two systems: The Event-based Text-mining of Health Electronic Records (ETHER) and the Pattern-based and Advanced Network Analyzer for Clinical Evaluation and Assessment (PANACEA). ETHER uses natural language processing to speed up and improve the review of the unstructured narrative portion of AE reports. PANACEA provides medical reviewers with advanced network analysis tools tailored to discover and investigate patterns among spontaneous reports. PANACEA and ETHER are tightly integrated, with capabilities for processing ETHER output in PANACEA and reviewing sets of cases from PANACEA in ETHER. As shown in the next sections, these tools may fill gaps related to the retrieval and analysis of structured and unstructured information and support certain safety surveillance and signal management processes.

In this manuscript, we describe the DSE and its use in post-market safety signal discovery and exploration. We first describe the DSE architecture and the methodologies underlying ETHER and PANACEA as well as the capabilities of each system. We next describe potential use cases for the DSE and illustrate these with an application to a spontaneous report dataset with a known safety concern. An accompanying video and an expanded supplement with screenshots from the two systems (see Supplement 1) illustrate the tight integration between PANACEA and ETHER, as well as key functionalities of both systems as applied to a real-world safety surveillance example.

2. Background

The identification and disposition of a safety signal is a complex process requiring medical judgment at every stage. Fig. 1 shows the processes and steps of the Signal Management Framework at the US FDA. Signals can arise from a variety of sources including passive surveillance (e.g., VAERS/FAERS), medical literature, clinical studies, periodic reports from manufacturers, observational studies/registries, and health claims data. MOs evaluate spontaneously reported adverse events by reviewing individual case source documents and/or encoded data. Additionally, MOs routinely perform product-specific data mining and literature reviews. Through applying their scientific, epidemiologic, and regulatory experience, MOs identify a safety signal when information from one or multiple sources suggests a new potentially causal association (or new aspect of a known association) judged to be sufficiently likely to warrant verification [8]. After detection, a signal is refined by various secondary processes including case series analyses, medical literature review/comparison, and (as needed) active surveillance. In a case series analysis, an expert reviews all available pertinent information (e.g., demographics, AE description, time to onset of AE, etc.) in order to identify any patterns [9]. MOs also conduct medical literature review to identify any pertinent information. Active surveillance may include retrieval of records from clinics or hospitals. These secondary processes of signal refinement help in signal characterization. If safety concerns persist, in collaboration with the management chain, the MOs formulate a plan for risk assessment (e.g., hypothesis testing in a pharmacoepidemiologic study) and risk communication to stakeholders for further regulatory action such as label change.

The Signal Management Framework contains a number of complex and demanding processes. As seen in Fig. 1, two of the primary processes are: the review of individual VAERS or FAERS reports and case series analyses. MOs are often challenged by the amount of information they have to review, particularly with the increasing number of annual submissions. In most cases, MOs need to read through the narratives to retrieve data that: (i) has not been coded, such as medical and family history; (ii) cannot be coded to MedDRA terms, such as temporal information or extensive hospital records; (iii) is coded incorrectly; or (iv) is not included in the structured fields, such as names of co-administered products that were not recorded at the time of submission. The critical steps of the Signal Management Framework cannot rely only on the listed MedDRA codes because these codes alone do not support the identification of certain key associations between the reported entities,

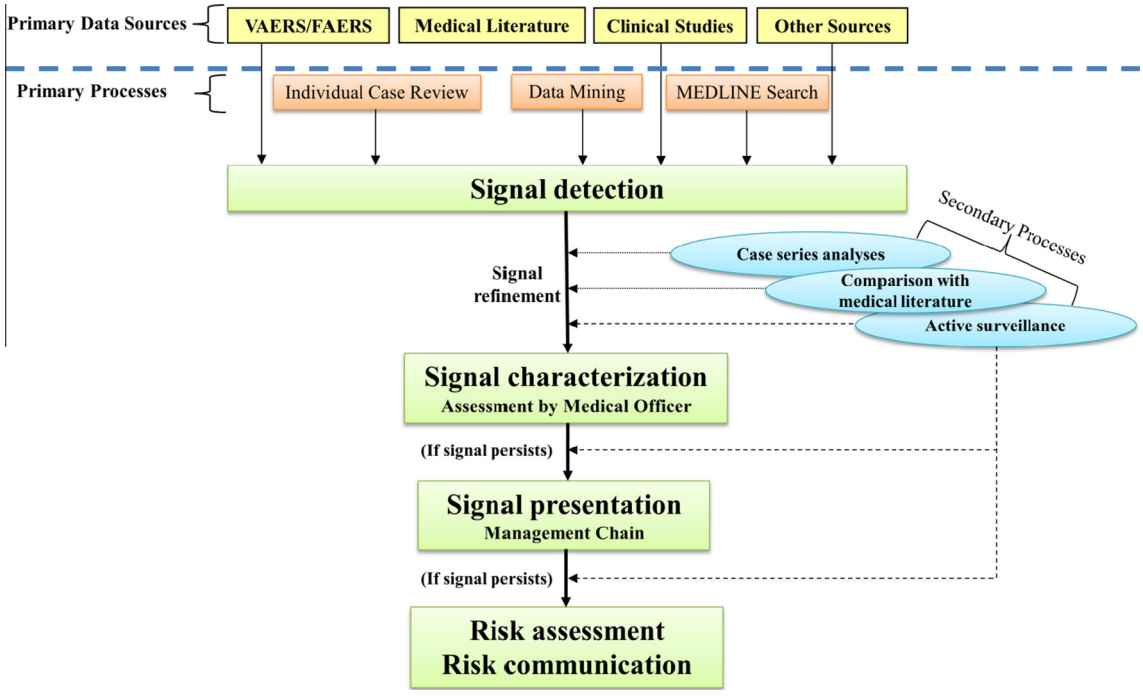

Fig. 1. The signal management framework.

Please cite this article in press as: T. Botsis et al., Decision support environment for medical product safety surveillance, J Biomed Inform (2016), http://dx.doi.org/10.1016/j.jbi.2016.07.023