Journal of Biomedical Informatics 76 (2017) 41–49

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Toward multimodal signal detection of adverse drug reactions

Rave Harpaz[a,*], William DuMouchel[a], Martijn Schuemie[b], Olivier Bodenreider[c], Carol Friedman[d], Eric Horvitz[e], Anna Ripple[c], Alfred Sorbello[f], Ryen W. White[e], Rainer Winnenburg[g], Nigam H. Shah[g]

[a] Oracle Health Sciences, Bedford, MA, United States
[b] Janssen Research and Development, Titusville, NJ, United States
[c] National Library of Medicine, NIH, Bethesda, MD, United States
[d] Columbia University, New York, NY, United States
[e] Microsoft Research, Redmond, WA, United States
[f] U.S. FDA, Silver Spring, MD, United States
[g] Stanford University, Stanford, CA, United States

ARTICLE INFO

Keywords:
Drug safety
Pharmacovigilance
Adverse drug reactions
Signal detection
Multimodal

ABSTRACT

Objective: Improving mechanisms to detect adverse drug reactions (ADRs) is key to strengthening post-marketing drug safety surveillance. Signal detection is presently unimodal, relying on a single information source. Multimodal signal detection is based on jointly analyzing multiple information sources. Building on, and expanding the work done in prior studies, the aim of the article is to further research on multimodal signal detection, explore its potential benefits, and propose methods for its construction and evaluation.

Material and methods: Four data sources are investigated; FDA's adverse event reporting system, insurance claims, the MEDLINE citation database, and the logs of major Web search engines. Published methods are used to generate and combine signals from each data source. Two distinct reference benchmarks corresponding to well-established and recently labeled ADRs respectively are used to evaluate the performance of multimodal signal detection in terms of area under the ROC curve (AUC) and lead-time-to-detection, with the latter relative to labeling revision dates.

Results: Limited to our reference benchmarks, multimodal signal detection provides AUC improvements ranging from 0.04 to 0.09 based on a widely used evaluation benchmark, and a comparative added lead-time of 7–22 months relative to labeling revision dates from a time-indexed benchmark.

Conclusions: The results support the notion that utilizing and jointly analyzing multiple data sources may lead to improved signal detection. Given certain data and benchmark limitations, the early stage of development, and the complexity of ADRs, it is currently not possible to make definitive statements about the ultimate utility of the concept. Continued development of multimodal signal detection requires a deeper understanding the data sources used, additional benchmarks, and further research on methods to generate and synthesize signals.

1. Introduction

The increasing harm and monetary burden associated with adverse drug reactions (ADRs) has made post-marketing drug safety surveillance (DSS) a top priority for health systems worldwide, and led to new legislative and research initiatives [1–6]. DSS presently relies on analysis of adverse event reports stored in spontaneous reporting systems (SRS) [7,8] such as the U.S. Food and Drug Administration (FDA) Adverse Event Reporting System (FAERS) [9,10] and the VigiBase maintained by the World Health Organization (WHO) [11] .

The complexity associated with numerous natural phenomena (such as the mechanisms, occurrence, or identification of ADRs) makes it is unlikely that a single information source provides complete knowledge of the phenomenon of interest. In many scientific domains, information about a given phenomenon can be acquired from multiple sources. Each such acquisition framework is referred to as a *modality* and is associated with one data source [12]. A system that provides access to information from multiple modalities is known as *multimodal*. The fusion and joint analysis of multiple modalities promises a more unified and global view of the problem at hand, and a solution that is greater than the sum of its parts.

The benefits that may come from multimodality have recently been

* Corresponding author at: 100 Crosby Drive, Bedford, MA 01730, United States.
E-mail address: rave.harpaz@oracle.com (R. Harpaz).

http://dx.doi.org/10.1016/j.jbi.2017.10.013
Received 14 January 2017; Received in revised form 14 October 2017; Accepted 24 October 2017

extended to DSS with the ability to process and analyze new kinds of observational, experimental, and knowledge-based data sources containing pertinent information about the effects, mechanisms, and safety of medical products. Among these data sources are: electronic health records and administrative claims [6,13–17], the biomedical literature [18–20], social media (e.g., health forums and social networks) [21–24], behavioral data drawn from the logs of search engines [25–27], and mechanistic information extracted from chemical and biological knowledge bases [28]. Together with SRS, each data source promises to provide a unique vantage point for shaping our understanding of a drug's safety profile, for contributing scientific evidence needed for causal assessment of adverse reactions, and for improving the timeliness of ADR identification.

In recent years, a large amount of research has been conducted on repurposing and analyzing safety information from such data sources, but relatively less progress has been made on development of multimodal approaches to jointly analyze the information provided by each data source.

A core component of DSS is ADR detection, and use of computational techniques known as signal detection are among its primary facilitators. These techniques enable drug safety researchers to analyze large volumes of data and generate hypotheses (signals) of new post-approval ADRs. Upon review, strong signals may lead to regulatory interventions such as drug withdrawals and the issuance of public warnings. From a technical standpoint, signal detection consists of applying computational methodologies to large databases in order to identify unexpected associations between drugs and potential adverse reactions [10].

Prior studies have demonstrated that pair-wise combinations of safety signals from several sources with FAERS can improve the accuracy of signal detection. Examples of signals that have been combined with those from FAERS include signals from EHRs [29,30], claims data [31], biomedical literature [32], chemical data [33], and Internet search logs [25]. Additional studies with less focus on signaling considered other forms of joint analysis of safety data [34,35].

The aim of this article is to strengthen empirical support for multimodal signal detection and provide additional insights on its performance characteristics. Building on, and extending the work performed in prior studies, we consider the joint analysis of more than just two data sources, and investigate a larger number of methods to synthesize signal statistics. Rather than relying on a single benchmark of well-known ADRs, we extend the evaluation with an additional benchmark of relatively new ADRs, and importantly, we incorporate the time aspect of signal detection into our performance evaluation. In doing so we propose new performance indices that blend both the accuracy of detection and the lead-time-to-detection.

We envision a multimodal signal detection system as one which pools and aggregates signal statistics from multiple data sources to produce a composite signal statistic (Fig. 1). In characterizing such a system, we examine four data sources, including three representative sources that would likely play a significant role in future systems, and a fourth novel, yet non-standard source that demonstrates promise. We use FDA's FAERS as a representative of SRS. As healthcare data we use a large database of administrative claims from millions of U.S. patients. The U.S. National Library of Medicine® (NLM) MEDLINE® citation database [36] serves as a source of biomedical literature, and health-related queries captured in the logs of major Web search engines are used as a source of consumer behavioral data.

Published methods are used to process and generate safety signals from each data source individually, and a range of recognized statistical approaches is investigated to transform signal statistics from each data source into a composite signal score. Performance is assessed through the use of both the retrospective and backdated prospective evaluation strategies, which are based on reference benchmarks made of well-established ADRs and recently labeled ADRs respectively.

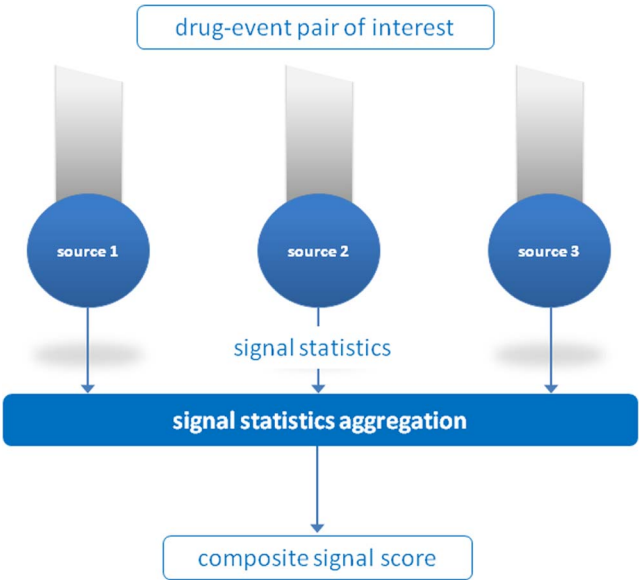

Fig. 1. Architecture of a multimodal signal detection system. Signal statistics are pooled and aggregated from different data sources to produce a composite signal statistic.

2. Material and methods

2.1. Data sources

The FAERS data consists of approximately six million reports collected by the FDA since the inception of the system (1968) to 2014Q2 [37]. The reports were preprocessed to correct reporting artifacts and were made available by Oracle Health Sciences.

Analysis of healthcare data was made possible through the Innovation in Medical Evidence Development and Surveillance Research Laboratory [38], which provides a secure computing environment, research tools, and five de-identified healthcare datasets for analysis. Of these datasets, we used the largest dataset called 'MarketScan Commercial Claims and Encounters', containing administrative claims from a privately insured U.S. population of approximately 142 million patients spanning the years 2003–2013 (inclusive).

Analysis of NLM's Medical Subject Headings (MeSH®) [39] descriptors was used to generate signals from MEDLINE bibliographic citations spanning the period from 1968 to May 2014. The algorithms for computationally identifying and extracting MEDLINE MeSH terms associated with ADRs are described by Winnenburg et al. [20]. The process resulted in approximately 360,000 ADR-related article citations and approximately 500,000 unique drug-event pairs potentially associated as ADRs, which were supplied along with other article metadata by the NLM.

De-identified search queries submitted to the Google, Bing, and Yahoo search engines by 80 million users during the period 2011–2013 were analyzed to generate signals based on behavioral data. This data source was accessed in collaboration with Microsoft Research, and described by White et al. [25]

2.2. Benchmarks

Our retrospective performance evaluation uses the Observational Medical Outcomes Partnership (OMOP) benchmark [40]. The OMOP benchmark consists of 399 unique drug-event pairs (test cases) of which 165 represent established true ADRs and 234 represent negative controls—pairs that are highly unlikely to be associated. Each pair is made of one of 181 drugs, and one of four serious adverse events (acute myocardial infarction, acute renal failure, acute liver injury, and upper