


Identifying Herbal Adverse Events From Spontaneous Reporting Systems Using Taxonomic Name Resolution Approach

Bioinformatics and Biology Insights
Volume 14: 1–11
© The Author(s) 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1177932220921350



Vivekanand Sharma¹ , Luiz Fernando Fracassi Gelin¹
and Indra Neil Sarkar^{1,2}

¹Center for Biomedical Informatics, Brown University, Providence, RI, USA. ²Rhode Island Quality Institute, Providence, RI, USA.

ABSTRACT: The efficacy and safety of herbal supplements suffer from challenges due to non-uniform representation of ingredient terms within biomedical and observational health data sources. The nature of how supplement data are reported within Spontaneous Reporting Systems (SRS) can limit analyses of supplement-associated adverse events due to the use of incorrect nomenclature or failing to identify herbs. This study aimed to extract, standardize, and summarize supplement-relevant reports from two SRSs: (1) Food and Drug Administration Adverse Event Reporting System (FAERS) and (2) Canada Vigilance Adverse Reaction (CVAR) database. A thesaurus of plant names was developed and integrated with a mapping and normalization approach that accommodated misspellings and variants. The reports gathered from FAERS between the years 2004 and 2016 show 185,915 herbal and 7,235,330 non-herbal accounting for 2.51%. The data from CVAR found 36,940 reports of herbal and 503,580 non-herbal reports between the years 1965 and 2017 for a total of 6.83%. Although not all cases were actual adverse events due to numerous variables and incomplete reporting, it is interesting to note that the herbs most frequently reported and significantly associated with adverse events were as follows: *Avena sativa* (Oats), *Cannabis sativa* (marijuana), *Digitalis purpurea* (foxglove), *Humulus lupulus* (hops), *Hypericum perforatum* (St John's Wort), *Paullinia cupana* (guarana), *Phleum pratense* (timothy-grass), *Silybum marianum* (milk thistle), *Taraxacum officinale* (Dandelion), and *Valeriana officinalis* (valerian). Using a scalable approach for mapping and resolution of herb names allowed data-driven exploration of potential adverse events from sources that have remained isolated in this specific area of research. The results from this study highlight several herb-associated safety issues providing motivation for subsequent in-depth analyses, including those that focus on the scope and severity of potential safety issues with supplement use.

KEYWORDS: Adverse Event Detection, Spontaneous Reporting Systems, herbal, indexing

RECEIVED: November 12, 2019. **ACCEPTED:** April 1, 2020.

TYPE: Original Research

FUNDING: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was funded in part by National Institutes of Health grants R01LM011963 and U54GM115677.

DECLARATION OF CONFLICTING INTERESTS: The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

CORRESPONDING AUTHOR: Indra Neil Sarkar, Center for Biomedical Informatics, Box G-R, Brown University, Providence, RI 02912, USA. Email: neil_sarkar@brown.edu

Introduction

Herbal supplements are used worldwide for reasons such as treating numerous ailments, performance enhancement, or for health maintenance. The estimated sale of herbal dietary supplements in the United States has reached \$7.452 billion in 2016 (an increase of 7.7% from the previous year).¹ There has been a consistent increase in herbal sales (over that last 13 years¹), and they are oftentimes perceived as safer alternatives to conventional drug interventions by consumers.² Although used by a large portion of the population, there are no guidelines in the United States for premarket approval of safety and efficacy for herbals, such as those provided by the US Food and Drug Administration (FDA) for drugs.³ Since most herbal intake generally falls outside the supervision of healthcare professionals, concerns of safety and efficacy are not routinely documented. An estimate of adverse events associated with dietary supplements (which includes herbals) suggests that only 1 in 100 are reported to FDA.⁴ In 2006, a bill (S.3546 “Dietary Supplement and Nonprescription Drug Consumer Act”; colloquially referred to as the “AER bill”) was signed into law (Public Law 109-462) in the United States Senate of the 109th Congress requiring manufacturers of dietary supplements (which includes botanical products) and

over-the-counter (OTC) products to submit serious adverse event reports (AERS) to FDA.⁵ Preliminary evidence among military personnel indicates that 22% of dietary supplement users have reported one or more adverse events.⁶ Dietary supplements that include herbals are sold on military bases in many countries.⁷ A committee on dietary supplement use by military personnel made recommendations about the critical requirement for identification of potentially harmful effects.⁸ The perception that herbal supplements (“botanicals”) are safer alternatives may not always be true (eg, St. John's Wort [*Hypericum perforatum*], a popular herbal antidepressant, has been shown to demonstrate adverse effects similar to the prescription medicine fluoxetine⁹). A manual assessment of case reports from PubMed revealed approximately 21, 7, and 7 herbs that posed risk of liver,¹⁰ kidney,¹¹ and heart¹² toxicities, respectively. There is limited information regarding potential adverse reactions associated with herbal treatments.² The lack of safety and efficacy information for herbs may play a role in hindering possible progress in identifying and assessing the safety of potential therapeutic candidates for alleviating symptoms associated with disease. There have been recommendations for including herbals in national pharmacovigilance systems.¹³ However, the lack of standardized labels can make



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

systematic analyses infeasible. For example, information in pharmacovigilance systems needs to include the identity of herbs instead of just brand names. The current pharmacovigilance framework requires much effort to systematically identify herbal or other supplement safety issues.

Identifying adverse drug reactions (ADR) have focused on data from Spontaneous Reporting Systems (SRS),^{14–16} biomedical literature,^{17–19} clinical reports (including Electronic Health Records),^{20–22} social media,^{23,24} and PubChem BioAssay.²⁵ Resources such as SIDER (Side Effect Resource²⁶) can provide comprehensive lists of drug-ADR information. In addition, the chemical, biological, and phenotypic properties of drugs have been used for larger-scale prediction of ADRs.²⁷ Publicly accessible SRS resources, such as the FDA Adverse Event Reporting System (FAERS)²⁸ and the Canada Vigilance Adverse Reaction (CVAR),²⁹ provide a means for monitoring safety of drug interventions in North America.^{2,30} However, the current data organization within these sources do not allow for systematic identification of AERs specific to herbals (eg, taxonomically). Lack of the ability for robust acquisition of relevant data therefore hinders subsequent design and execution of data-driven analysis of associated health outcomes.

The compilation and detection of potential adverse signals for herbal ADRs from SRS is challenging. This is in part due to the variability of their inclusion in public SRSs (since they are required to be reported only for serious adverse events that result in hospitalization, significant disability, or death⁵). Herbal interventions can be reported in SRS as a range of brand names, vernacular names, and scientific name synonyms, often embedded within free text without association with a standardized code or nomenclature. Taking full advantage of SRS data therefore requires data cleaning, mapping, and normalization to controlled terminologies. For pharmaceutical drugs, several efforts have been made in standardization and curation. Natural language processing tools such as MedEx³¹ or other drug mapping tools^{30,32} have been used to normalize reported drug names to RxNorm.³³ Community efforts such as the Observational Health Data Sciences and Informatics (OHDSI) provide a platform for researchers to gain access to standardized data for generating insights through analytics.³⁴ Resources such as the Adverse Event Open Learning through Universal Standardization (AEOLUS) facilitate interoperability and ease of analysis by making available FAERS data in a standardized form.³⁰ The biomedical domain is equipped with standard vocabularies from several sources that allow such standardization. The existing gap between dietary supplement terms and standard terminologies limits the ability to retrieve associated health outcome data.³⁵ To date, there has been limited direct use of such approaches for studying events associated with herbal interventions from public SRS data. Amid the acknowledged importance of identifying herbal-specific adverse events, to date there has been limited use of informatics approaches to systematically identify herbal-specific AERs and

consequential signals that can be used to motivate subsequent safety studies.

This study explored the feasibility of extracting herbal adverse events from two SRSs, specifically, FAERS and CVAR. This resource was integrated with a mapping and alignment approach accommodating spelling errors and approximate matches to identify records related to herbals. The results demonstrate that there is indeed a significant amount of herb-specific information embedded within SRSs that can be used to analyze associated botanical adverse events.

Materials and Methods

The main goal of this study was to develop an approach to facilitate the acquisition of herb-related adverse event information from SRS. A thesaurus of plant names was created by combining entries from three major taxonomic sources that further grouped scientific names, synonyms, and vernaculars. Using this as a nomenclature scaffold, a mapping approach was designed to accommodate detection of variants and spelling errors. The intervention name strings from SRS were processed to identify and resolve mappings to standardized taxonomic plant names. The standardized data were used to calculate statistically significant safety signals at two different levels of granularity, leveraging the hierarchical structure of the Medical Dictionary for Regulatory Activities (MedDRA). A general overview of the approach is graphically depicted in Figure 1.

Data sources used for the study

Data from two SRS were considered for this study: (1) FAERS and (2) CVAR database.

FAERS. FAERS contains information on adverse event and medication error reports that are submitted to the FDA as a part of post-marketing surveillance program for drug and therapeutic biologic products. Like CVAR, these reports come from consumers, health providers, and manufacturers. The reports in FAERS covering data from 2004 to 2016 were used in this study. This dataset consisting of seven tables was loaded into MySQL database, accommodating for changes in format for records before August 2012. Relevant information related to interventions, active ingredients, demographics, and adverse events were collected for further processing. De-duplication of FAERS records was carried out by FDA's recommendation. For reports with the same CASE number, the most recent report (FDA_DT) was retained and when the CASE and FDA_DT values were the same, the highest ISR number was selected.

CVAR. CVAR is a Canadian online database that contains reports of adverse reactions, including post-marketing ADR data that have been reported by consumers and health professionals. In addition to this, reports from manufacturers and distributors are also included. This database includes reports

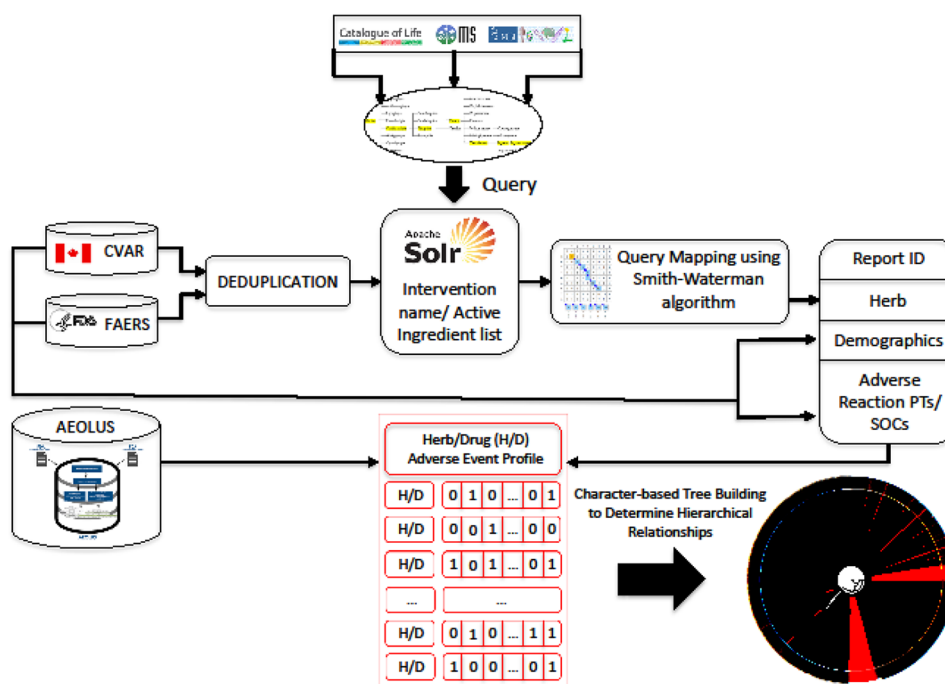


Figure 1. Overview of the approach followed in this study.

AEOLUS: Adverse Event Open Learning through Universal Standardization; CVAR: Canada Vigilance Adverse Reaction; FAERS: FDA's Adverse Event Reporting System; PT: Preferred Term; SOC: System Organ Class.

starting from 1965. The ADR data obtained from CVAR at the time of this study covered reports up to March 31, 2017. The raw data consisting of 11 flat files that were loaded into a local MySQL relational database. From these data tables, information related to interventions, ingredient, demographics, and adverse reactions were collected for further processing. The linked/duplicate reports information from the Report_Links_LX table was used for de-duplication of CVAR records.

Identification of records associated with herbals

The pipeline for identification of herb-associated AERS involved two steps: (1) creating an herb-specific thesaurus and (2) mapping and resolution of intervention name strings to herb names.

Plant species thesaurus (uBiota). A unified compendium of plant species names, synonyms, and vernaculars ("uBiota") was created from the union of three sources. The data from the sources were organized by accepted scientific names to which unique identifiers were assigned keeping track of source and source identifiers. Synonyms and vernacular names were also isolated, relations labeled, and organized based on unique identifiers. To maintain the tractability and feasibility, uBiota represents the seven canonical Linnaean taxonomic groups: Kingdom, Phylum, Class, Order, Family, Genus, and Species. A short description of the three sources chosen for this study is provided:

1. Catalog of Life (COL): This is the most comprehensive and authoritative global index of species compiled from

diverse sources around the world. It provides critical species information on synonymy, higher taxa, and distribution of across global regions.

2. Integrated Taxonomic Information System (ITIS): As a result of recognition of the importance of organization and access to standardized nomenclature by the White House Subcommittee on Biodiversity and Ecosystem Dynamics, this resource was designed. The database ensures high quality with valid classifications, revisions, and additions of newly described species.
3. NCBI Taxonomy: This resource provides manually curated classification and nomenclature for all the organisms in the public sequence databases.

Mapping of intervention name strings. The mapping and normalization involved indexing the intervention name string or active ingredient name string dataset from FAERS and CVAR using Apache Solr. The retrieval step leveraged a fuzzy-matching feature provided by Solr. The retrieved strings were subjected to further verification using Smith-Waterman string alignment algorithm. Based on the alignment, matches were determined using the following constraints: (1) The coverage of aligned string matched the length of taxonomic name, (2) the difference between herb name length and alignment score is less than half of herb name length, (3) the first and last characters of herb name matched, and (4) if the length of herb name was less than or equal to four characters, then, perfect matches were considered. The identified herb names were then resolved to their respective scientific names. A more detailed description, evaluation, and a comparison of this mapping

approach with other existing tools is provided in our previously published study.³⁶

Evaluation of intervention name mapping. The evaluation of mapping of intervention name strings was carried out on 1000 uncorrected name strings from SALVIAS database (provided and used previously by Boyle et al³⁷). Assessment was based on standard evaluation metrics of Precision, Recall, and F-score.

Detection of adverse event signals

Signal Disproportionality Analysis (SDA) was conducted at the level of Preferred Term (PT) and System Organ Class (SOC) from MedDRA hierarchy. Two metrics for signal detection were used: (1) the Proportional Reporting Ratio (PRR) and (2) Reporting Odds Ratio (ROR). The significance of the signal was determined as described by Gavali et al.³⁸ (1) If the number of co-occurrences were three or more, (2) the PRR or ROR value was greater than or equal to three, and (3) the lower bound of 95% confidence interval (CI) greater than or equal to one. For the purpose of reporting, a filter was set to retain only the most reliable adverse events associations. This filter accommodated the confounding factor of herb name variations, allowed only those associations where the scientific name was present at least once.

Herb-drug adverse event similarity analysis

The significant associations identified using SDAs from the two data sources (FAERS and CVAR) were merged. To compare herbal adverse event profiles with those of drugs, this dataset was combined with significant associations from AEOLUS.³⁰ AEOLUS is a publicly available standardized resource that provides a pre-processed and clean version of FAERS data along with pre-computed statistical measures of disproportionality (PRR and ROR). The combined dataset was used for hierarchically grouping of herbs/drugs using the maximum parsimony phylogenetic technique. The input matrix for the phylogenetic analysis was constructed such that adverse event terms were as represented as characters and their respective presence ("1") or absence ("0") as character states. These adverse event profile data were then analyzed using the maximum parsimony phylogenetic inference tool, tree analysis using new technology (TNT).³⁹ The output tree was evaluated using the drug pharmacodynamic pathway entries listed in PharmGKB.⁴⁰ For each selected pathway, related pathways were gathered and compared against drug group subsets from the generated tree. The evaluation was based on whether drugs belonging to the related pathways were grouped. A web tool called PhyloZoom was used to enable interactive exploring of the tree (available at <https://bcbi.brown.edu/phylozoom/>). Selected evaluation based on comparison with results from Mizutani et al⁴¹ is also presented in Figure 3A to C.

Table 1. AER counts from CVAR and FAERS.

SOURCE	HERBAL	NON-HERBAL	TOTAL	%
FAERS	185,915	7,235,330	7,421,245	2.51
CVAR	36,940	503,580	540,520	6.83

AER: adverse event report; CVAR: Canada Vigilance Adverse Reaction; FAERS: FDA's Adverse Event Reporting System. Timeframe: FAERS:2004-2016 and CVAR: 1965-2017.

Results

Summary of herb thesaurus and mapping evaluation

The compilation of plant species names and their taxonomy from COL, ITIS, and NCBI taxonomy resulted in 390,638 scientific names, 660,645 synonyms, and 82,672 vernacular names. The evaluation of the pipeline on SALVIAS dataset resulted in Precision, Recall, and *F*-score of 0.92, 0.87, and 0.92, respectively.

Summary of herb-related reports

The identified herb-related reports from FAERS accounted for 2.51% of reports. Herb-related reports from CVAR accounted for 6.83% of total reports (Table 1). In general, it was observed that AERs (both herb and non-herb) were higher for females. However, the proportions (group specific herb-AERs out of total herb-AERs) of herb-AERs were similar between male and female groups for both FAERS and CVAR dataset. The proportion of non-herb AERs was similar or higher in all age groups below 65. For the age-group 65 and higher, the proportion of herb-AERs was significantly higher when compared to non-herb AERs (*P*-values calculated using 2-sample test for equality of proportions with continuity correction) (Figure 2).

Herb-associated adverse event signals

The identified adverse event associations from FAERS dataset accounted for 109 and 130 plant species associated with PTs and SOCs, respectively. Similarly, from CVAR dataset comprised of 298 and 249 plant species associated with PTs and SOCs, respectively. The adverse event counts from FAERS and CVAR are provided in Table 2. Table 2 also provides PRR values and CIs of those associations that were reported at least once as "PS/SS" (FAERS dataset) or "Suspect" (CVAR dataset). The full set of identified associations from both the data sources along with their respective SDA scores are provided as supplemental tables. A small subset of identified adverse events that were found to be in common between FAERS and CVAR and significant with at least 10 or more reports is provided in Tables 3 and 4.

Herb-drug adverse event similarity

In order to determine the severity and extent of adverse events associated with herbs as well as their respective potential to cause harm a comparison with prescription drugs was carried

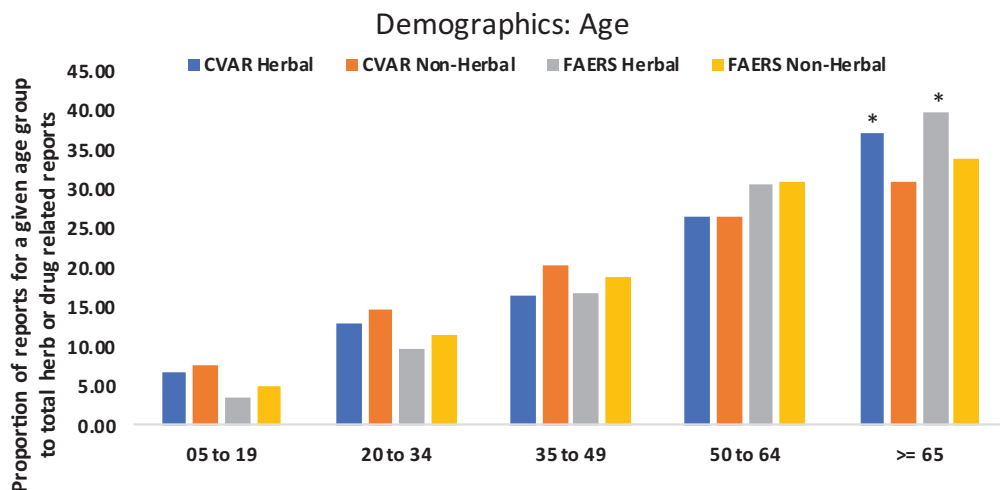


Figure 2. AER count comparison among different age groups (Asterisks indicate significant difference). Timeframe: CVAR: 1965-2017 and FAERS:2004-2016. AER: adverse event report; CVAR: Canada Vigilance Adverse Reaction; FAERS: FDA’s Adverse Event Reporting System.

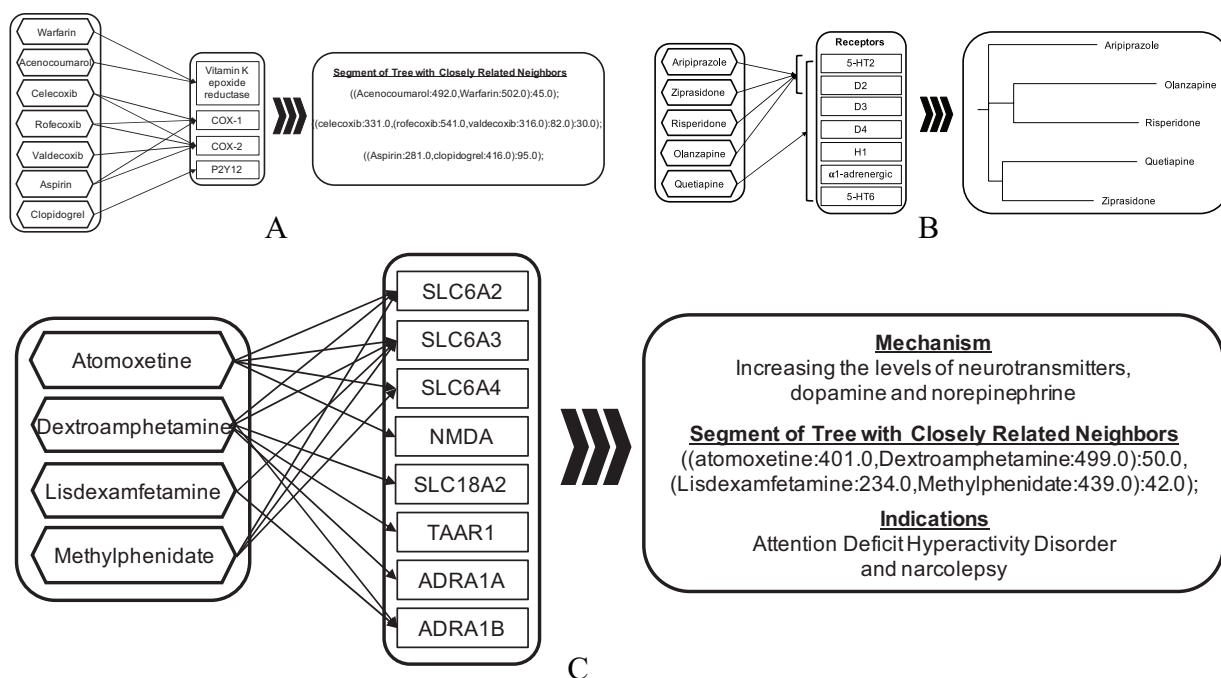


Figure 3. Evaluation of hierarchical grouping of herbs/drugs using the maximum parsimony phylogenetic technique. Groupings of drugs (e.g., A, B, and C represent three different groupings) with known underlying mechanisms were manually examined to validate the efficacy of the approach used (evaluation based on comparison with results from Mizutani et al).

Table 2. Counts of adverse events and System Organ Class (SOC) that were found significant based on both PRR and ROR values and their respective Confidence Intervals.

SOURCE	ADVERSE EVENTS		SOC	
	ALL	SUSPECT	ALL	SUSPECT
FAERS	3437	883	635	326
CVAR	3354	2600	390	340
Common	248	52	45	14

CVAR: Canada Vigilance Adverse Reaction; FAERS: FDA’s Adverse Event Reporting System; PRR: Proportional Reporting Ratio; ROR: Reporting Odds Ratio; SOC: System Organ Class. Timeframe: FAERS:2004-2016 and CVAR: 1965-2017.

out. The evaluation of hierarchical relationships among drugs/herbs based on their respective adverse event profiles using the maximum parsimony criterion revealed characteristically distinguishable groupings. Selected examples are shown in Figure 3A to C as evaluation of validity of the approach. Drugs with similar mechanisms of action and indications were grouped together. Warfarin and Acenocoumarol, anticoagulants targeting Vitamin K epoxide reductase, were discovered in the same cluster. Nonsteroidal anti-inflammatory drugs Celecoxib, Rofecoxib, and Valdecoxib (targeting COX-1 and COX-2) were also grouped together as having similar adverse events. In addition, Aspirin and Clopidogrel, although having different targets but similar indications of use and adverse reactions,⁴¹

Table 3. Common System Organ Class (SOC)^a from CVAR and FAERS.

HERB	SOC	FAERS		CVAR	
		PRR	CI	PRR	CI
<i>Avena sativa</i> (Oats)	Skin and subcutaneous tissue disorders	9.32	3.67-23.67	3.87	3.63-4.12
<i>Cannabis sativa</i> (Marijuana)	Psychiatric disorders	4.77	4.47-5.09	5.17	4.87-5.48
	Social circumstances	8.2	7.37-9.12	7.96	6.11-10.36
<i>Digitalis purpurea</i> (Foxglove)	Cardiac disorders	8.5	3.3-21.93	4.72	3.4-6.56
<i>Humulus lupulus</i> (Hops)	Psychiatric disorders	6.53	3.59-11.89	3.66	2.42-5.53
<i>Hypericum perforatum</i> (St John's Wort)	Psychiatric disorders	4.21	3.64-4.86	3.43	2.74-4.29
<i>Paullinia cupana</i> (Guarana)	Cardiac disorders	3.54	1.99-6.31	4.72	3.52-6.32
<i>Peumus boldus</i> (Boldo)	Investigations	8.56	3.44-21.28	4.12	2.69-6.31
<i>Phleum pretense</i> (Timothy-grass)	Immune system disorders	13.27	8.03-21.92	12.47	7.91-19.66
<i>Silybum marianum</i> (Milk thistle)	Hepatobiliary disorders	3.45	2.95-4.04	4.57	2.94-7.12
<i>Taraxacum officinale</i> (Dandelion)	Hepatobiliary disorders	3.67	2.02-6.67	5.03	2.79-9.08
<i>Valeriana officinalis</i> (Valerian)	Psychiatric disorders	3.31	2.18-5.02	3.29	2.48-4.36

CI: confidence interval; CVAR: Canada Vigilance Adverse Reaction; FAERS: FDA's Adverse Event Reporting System; PRR: Proportional Reporting Ratio; SOC: System Organ Class. Timeframe: FAERS:2004-2016 and CVAR: 1965-2017.

^aNot all data represent 100% adverse events; PRR scores are provided alongwith confidence intervals. ROR scores are available in Supplementary Tables.

Table 4. Common adverse event associations of herbs found in FAERS and CVAR.

HERB	ADVERSE EVENT	PRR [CONFIDENCE INTERVAL]
<i>Allium sativum</i> (Garlic)	Blood pressure increased	3.09 [2.45-3.89]; 3.34 [1.99-5.61]
<i>Camellia sinensis</i> (Tea)	Heart rate increased	3.21 [2.01-5.12]; 4.65 [2.87-7.55]
<i>Cannabis sativa</i> (Marijuana)	Suicide attempt	4.30 [3.30-5.61]; 5.37 [2.97-9.70]
	Aggression	7.19 [5.85-8.84]; 9.48 [6.78-13.25]
	Loss of consciousness	3.58 [2.94-4.36]; 3.76 [2.55-5.54]
	Psychotic disorder	6.31 [4.69-8.49]; 5.84 [3.52-9.70]
	Suicidal ideation	4.78 [3.91-5.84]; 4.25 [2.73-6.61]
	Euphoric mood	27.44 [21.58-34.89]; 44.46 [33.34-59.28]
	Dysarthria	3.11 [2.13-4.54]; 5.92 [3.36-10.43]
	Overdose	5.59 [4.76-6.56]; 9.99 [7.15-13.96]
	Toxicity to various agents	8.81 [7.55-10.28]; 4.20 [2.32-7.59]
	Anger	4.34 [3.13-6.02]; 7.77 [4.41-13.67]
	Drug withdrawal syndrome	5.43 [4.41-6.69]; 29.68 [22.24-39.61]
	Drug diversion	45.72 [32.23-64.86]; 122.31 [89.85-166.50]
<i>Curcuma longa</i> (Turmeric)	Fatigue	3.36 [2.93-3.85]; 3.38 [2.04-5.59]
<i>Glycine max</i> (Soybean)	Alanine aminotransferase increased	9.03 [5.90-13.82]; 3.18 [1.76-5.76]
	Aspartate aminotransferase increased	7.68 [4.74-12.44]; 3.75 [2.17-6.48]
<i>Hypericum perforatum</i> (St John'sWort)	Paraesthesia	8.08 [6.14-10.63]; 6.56 [3.55-12.13]
	Tremor	5.79 [4.26-7.88]; 6.30 [3.49-11.38]

(Continued)

Table 4. (Continued)

HERB	ADVERSE EVENT	PRR [CONFIDENCE INTERVAL]
	Anxiety	7.42 [5.91-9.31]; 10.37 [6.17-17.44]
	Depression	8.01 [6.39-10.04]; 7.07 [3.72-13.44]
<i>Linum usitatissimum (Flax)</i>	Joint swelling	3.77 [2.66-5.35]; 7.60 [4.15-13.92]
<i>Medicago sativa (Alfalfa)</i>	Diarrhea	3.80 [2.23-6.48]; 5.03 [3.00-8.43]
	Asthenia	5.35 [3.14-9.13]; 5.26 [2.89-9.57]
<i>Senna alexandrina (Senna)</i>	Confusional state	3.85 [3.53-4.20]; 3.23 [2.78-3.75]
	Acute kidney injury	6.28 [5.53-7.13]; 3.49 [2.69-4.53]
	Hypokalemia	5.19 [4.51-5.97]; 3.76 [2.57-5.50]
	Urinary retention	3.85 [3.18-4.66]; 3.50 [2.44-5.02]
	Constipation	5.23 [4.86-5.62]; 4.46 [3.82-5.20]
	Respiratory failure	3.47 [3.05-3.95]; 3.49 [2.35-5.18]
	Pancytopenia	4.10 [3.56-4.73]; 3.57 [2.46-5.18]
	Neutropenia	4.48 [4.04-4.97]; 3.29 [2.66-4.07]
	Delirium	4.50 [3.77-5.37]; 5.48 [4.16-7.22]
	Hyperkalemia	5.22 [4.45-6.12]; 3.57 [2.26-5.63]
	Blood creatinine increased	3.64 [3.19-4.16]; 3.77 [3.09-4.60]
	Red blood cell count decreased	4.07 [3.30-5.01]; 4.73 [3.36-6.65]
	Depressed level of consciousness	3.45 [2.88-4.13]; 4.36 [3.25-5.86]
	Pneumonia aspiration	5.38 [4.44-6.51]; 11.08 [7.91-15.52]
	International normalized ratio increased	4.67 [4.05-5.39]; 4.58 [3.29-6.38]
	Ascites	3.55 [2.89-4.35]; 3.96 [2.60-6.04]
	Plasma cell myeloma	5.02 [4.14-6.08]; 5.85 [3.89-8.81]
	Platelet count decreased	4.19 [3.76-4.67]; 3.41 [2.59-4.49]
	Haematocrit decreased	6.32 [5.35-7.46]; 6.39 [4.36-9.37]
	Neutrophil count decreased	5.01 [4.23-5.93]; 3.65 [2.62-5.09]
	Hypophagia	4.64 [3.78-5.70]; 3.77 [2.33-6.11]
<i>Taraxacum officinale (Dandelion)</i>	Dizziness	4.59 [2.63-8.02]; 3.22 [1.77-5.86]
<i>Vaccinium macrocarpon (Cranberry)</i>	Constipation	3.09 [2.57-3.71]; 7.71 [4.49-13.23]
	Sedation	3.59 [2.29-5.63]; 38.29 [20.41-71.84]
	Urinary tract infection	10.43 [9.29-11.70]; 7.59 [4.14-13.90]
	Fall	3.63 [3.19-4.13]; 6.04 [3.73-9.77]
	Cognitive disorder	3.64 [2.60-5.10]; 40.27 [21.47-75.54]
<i>Vaccinium myrtillus (European Blueberry)</i>	Vomiting	4.77 [3.32-6.85]; 3.59 [1.96-6.56]
<i>Zingiber officinale (Ginger)</i>	Nausea	4.40 [3.67-5.27]; 3.31 [2.32-4.73]
	Tremor	3.30 [2.26-4.81]; 3.53 [1.93-6.47]

CVAR: Canada Vigilance Adverse Reaction; FAERS: FDA's Adverse Event Reporting System; PRR: Proportional Reporting Ratio; ROR: Reporting Odds Ratio; Timeframe: FAERS: 2004-2016 and CVAR: 1965-2017.

PRR scores are provided along with confidence intervals from FAERS and CVAR in order. ROR scores are available in Supplementary Tables.

Note: Not all data represent 100% adverse events.

Table 5. Herb-Drug pairs placed closest to each other by hierarchical grouping based on tree analysis.

HERB ID	HERB NAME	DRUG ID	DRUG NAME	DISTANCE
pt_1729664	<i>Anamirta cocculus</i>	dr_19132970	Rose Bengal Sodium I 131	4
pt_846617	<i>Salvia miltiorrhiza</i>	dr_19059658	Chromocarb diethylamine	4
pt_833015	<i>Aralia elata</i>	dr_40125462	Chlorpheniramine/ibuprofen/pseudoephedrine oral suspension	6
pt_1780721	<i>Commiphora wightii</i>	dr_19058613	Isopropamide	8
pt_696557	<i>Melilotus officinalis</i>	dr_45892826	Potassium oxide	10
pt_1852918	<i>Angelica archangelica</i>	dr_42705857	Linagliptin/metformin oral tablet	13
pt_696725	<i>Myroxylon balsamum</i>	dr_42899780	Sesame oil	26
pt_820366	<i>Picea abies</i>	dr_40173337	Polysorbate 20	28
pt_1796298	<i>Viscum album</i>	dr_42900022	Viscum album leaf extract	29
pt_934627	<i>Hedera helix</i>	dr_40039679	Furosemide/potassium extended release oral tablet	32
pt_1842602	<i>Nicotiana tabacum</i>	dr_19018431	Alminoprofen	35
pt_693531	<i>Arachis hypogaea</i>	dr_44814342	Menthol/zinc oxide topical lotion	37
pt_1883838	<i>Echinacea purpurea</i>	dr_19029278	Diethylamine salicylate	38
pt_1725444	<i>Peumus boldus</i>	dr_40132089	Omeprazole/sodium bicarbonate oral suspension	91
pt_1828044	<i>Plantago asiatica</i>	dr_19097405	Fusafungin	97
pt_883917	<i>Convallaria majalis</i>	dr_43013119	Lomitapide	103
pt_1741204	<i>Vitis vinifera</i>	dr_43525901	Vitis vinifera leaf extract	107
pt_1764634	<i>Crataegus laevigata</i>	dr_43013119	Lomitapide	118
pt_1842772	<i>Atropa belladonna</i>	dr_1116109	Belladonna Alkaloids	162
pt_882469	<i>Aloe vera</i>	dr_19016670	Docosahexaenoate	198
pt_703486	<i>Glycine max</i>	dr_19040627	Sultamicillin	255
pt_1766223	<i>Cannabis sativa</i>	dr_40009527	Amphetamine/dextroamphetamine oral tablet	375
pt_849191	<i>Serenoa repens</i>	dr_1391307	Saw palmetto extract	387
pt_888989	<i>Ginkgo biloba</i>	dr_1391889	Ginkgo biloba extract	517
pt_693065	<i>Senna alexandrina</i>	dr_1337651	Vorinostat	652

were paired together. Similarly, based on related ADRs Aripiprazole, Ziprasidone, Risperidone, Olanzapine, and Quetiapine were grouped together, as well as validated by previous work.⁴¹ Drugs indicated for attention deficit hyperactivity disorder and narcolepsy showed clear grouping with related mechanism of increasing levels of neurotransmitters (dopamine and norepinephrine). Results from more thorough investigations of grouping of drugs are presented in the Supplementary data. The closest neighbors (drug concepts from AEOLUS) associated with herbal interventions identified in this study are listed in Table 5, where distance is defined by the parsimony criterion. Five out of the 25 identified closest associations exhibit same herb concepts (from this study and from AEOLUS) paired together further validating the efficacy of approach used in grouping herbs/drugs based on similarity

of adverse event profiles. The associations in Table 5 show several herbs that have adverse event profiles similar to prescription drugs.

Discussion

Although widely used, the issues of efficacy and safety of botanicals still remain a concern. There have been studies focused on evaluating such aspects of botanicals, but the results are isolated and embedded within large amounts of biomedical literature.⁴² Available studies of ADRs are represented in the form of systematic reviews.² In addition to these challenges, most cases of herb-related ADRs remain under-reported in post-marketing SRS.

Apart from these obstacles, data-driven studies to gain insights by analyzing AERs from multiple data sources are

notably challenged by the lack of being organized according to data organizing standards.³⁵ Resolving the mentions of supplement names and ingredients to their respective plant species names is a primary requirement for further identification and analysis of reports. The interventions can be mentioned by vernacular names, synonyms, scientific names, or product names. There is a clear lack of standardization of intervention names that is an important criterion for facilitating the identification of herb specific ADRs. Standardization and resolution of herb mentions as well as adverse event terminology across databases such as LiverTox⁴³ and Dietary Supplement Label Database⁴⁴ may possibly enhance connectivity providing avenues to generate data-driven insights. The associations identified here may be a great resource to enrich the content of existing toxicity databases but will require manual assessment. Leads from analysis of data across several sources has the potential to design more informative labels for available herbal supplements in the market thereby helping consumers make informed choices.

The goal of this study was to test the feasibility of mapping and extracting herbal ADRs from two of the current SRSs, FAERS, and CVAR. Resolving herb species name was a major issue. After a preliminary examination of the intervention names listed in these two databases, it was identified that these data sources required an approach for effective mapping and resolution of herb names that could be scaled to accommodate the characteristically larger size. Ambiguity and synonymy of taxonomic names poses a major issue in the study of biodiversity. There is still a pressing need for automated systems that recognize taxonomic names overcoming the aforementioned hurdles.⁴⁵ Lack of such resources that provide authoritative taxonomic solutions hinder the process of integration, and thereby interoperability of knowledge, between biodiversity and biomedical data sources.⁴⁶

AERs were identified for herbal interventions from both FAERS and CVAR. Based on a comparison of demographic variables (sex and age) for herb versus drug-related reports, a difference in the proportion of associated AERs was observed. This is consistent with previous estimates that older adults are more likely to use supplements.³ The AER data extracted from FAERS and CVAR was used for detection of adverse event signals at the level of associated adverse events and SOC. ADRs at the SOC level were calculated as they are known to improve the detectability of signals.²⁵ Significant associations from both FAERS and CVAR were identified (Supplementary Data). A list of significant ADRs that were common among FAERS and CVAR is provided in Tables 3 and 4. Some of the herbs identified previously in an assessment of PubMed displayed significant adverse effect associations. For example, germander, black cohosh, kava kava, and green tea with liver toxicity.¹⁰ Kava kava was also associated with renal and urinary disorders (SOC). Similarly, St John's Wort shows kidney toxicity¹¹ as identified from FAERS dataset. Some of the other high scoring herb-SOC associations common among the two datasets were as follows: (1) Timothy-grass, which is a known allergen but it's

extract is used to help the body develop immunity; (2) St John's Wort, marijuana, hops, and valerian that are known for their psychotropic properties were associated with psychiatric disorders; and (3) Dandelion and milk thistle were associated with hepatobiliary disorders. These herbs are commonly known to improve liver function. Having them associated with liver toxicity may be a result of discrepancy in reporting; (4) Foxglove and guarana showed adverse effects within the SOC category of cardiac disorders. Although, foxglove is used for heart related issues, the toxicity may arise as a response to higher dose. Guarana, a common ingredient in energy drinks has been known to display cardiovascular adverse effects when taken in larger quantity.⁴⁷ In addition to the signals identified in this study, there is a possibility for identification of more ADRs from reports that may appear non-significant as a result of the underreported nature of herbal AERs.

A unified compendium of taxonomic names, synonyms, and vernacular names was created and used along with a mapping strategy for quick and accurate mapping allowing for identification of misspelled names.³⁶ Such unification of taxonomic names from three sources provides a more comprehensive coverage for resolving names that would be missed out otherwise leading to incomplete information recovery. This resource is available for use from <https://bcbi.brown.edu/solrplant/>. At present, this tool doesn't resolve authority from scientific names and will be added in the future.

The challenges in normalizing herbal ingredient names and the underreported nature of herbal ADR data are major hurdles in identifying adverse reaction associated with herbals. However, additional challenges may exist in mining SRS data as a result of evolving PTs in MedDRA versions that are used for coding ADRs.³² MedDRA versions are updated twice annually and some codes may become obsolete in more recent versions. Such changes can result in missing reports during the identification of AERs, which in turn may lead to missing adverse event signals. To overcome this challenge, future work will involve mapping of listed adverse event PTs to a standardized clinical terminology resource such as Systematized Nomenclature of Medicine—Clinical Terms (SNOMED-CT)⁴⁸ or Adverse Drug Reaction Classification System (ADReCS).⁴⁹ A cursory survey of the data also revealed that the herbal supplements listed in the SRS are taken in combination with one or more other drugs or supplements. This suggests that the observed ADRs may also be a result of interactions with other drugs or supplements in addition to the direct effects. There is also a possible confounding factor that sex, age, and underlying medical conditions may be the cause of supplement intake related ADR. Although important, this area has received very little attention as reflected from the significant shortage of scientific literature and warrants in-depth investigation.

To gain a comparative view of ADR profiles of herbs and drugs, this study analyzed the grouping and categorization using a maximum parsimony tree-building framework. The use of a character-based approach enables the identification of

discrete characters shared among entities.⁵⁰ Inspection of the resulting tree for drugs with known indications and mechanism of action revealed meaningful groupings. Drugs with common underlying mechanisms may result in similar ADRs. Such associations have been studied using the drug indication information and their associated ADRs. The groupings were evaluated based on this premise, and the results were compared to related pathways from PharmGKB. Selected examples are provided in Figure 3A to C that were used as evaluation by comparing the results discussed by Mizutani et al⁴¹ Representative closest drug neighbors associated with herbs were also extracted (presented in Table 5). Herbs placed in proximity to pharmaceutical drugs have similar adverse event profile which may have remained unnoticed. For example, the hierarchical clustering shows comparable adverse event profiles as well as potential interaction⁵¹ for *Cannabis sativa* and amphetamine/ dextroamphetamine. In addition, such data may shed some light into associated or similar mechanisms underlying ADRs and potential medicinal use. However, few such associations were uncovered from the analysis done in this study, suggesting the possibility that either use of herbals may not result in as severe ADRs as prescription drugs or there may not be enough reports to make conclusive arguments. Future work will include the analysis of signals arising from rare events to highlight potential ADRs of interest. The results of this study suggest that there are notable herbal ADRs that the existing SRS infrastructure is well poised to support monitoring of botanical use and support clinicians in understanding the potential risk factors faced by their patients who choose to include herbals in their regimen. Such monitoring becomes more essential when used in combination with prescription drugs justifying in-depth data-driven investigation of herb–drug interactions.

Conclusion

Evidence from observational and biomedical data sources have the potential to provide significant insight into efficacy and safety issues related to currently marketed dietary and herbal supplements. This study focused specifically on identifying herbals from SRSs using a combination of informatics approaches. The promising results suggest that systematic approaches involving appropriate mapping and normalization of herb names may facilitate interoperability among traditionally disconnected biodiversity, biomedical, and observational data sources.

Authors' Note


The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions

VS and INS conceptualized and designed the overall study. VS developed the systems and approach to gather and analyze the

data. FG developed the tree-based analytic tool and implemented portions of the conceptualized pipeline. VS and INS interpreted the results. All authors contributed to, drafted, and reviewed the manuscript.

ORCID iD

Vivekanand Sharma  <https://orcid.org/0000-0003-1188-2312>

Supplemental Material

Supplemental material for this article is available online.

REFERENCES

- Smith T, Kawa K, Eckl V, Morton C, Stredneyd R. Herbal supplement sales in US increase 7.7% in 2016. Consumer preferences shifting toward ingredients with general wellness benefits, driving growth of adaptogens and digestive health products. *HerbalGram*. 2017;115:56–65.
- Izzo AA, Hoon-Kim S, Radhakrishnan R, Williamson EM. A critical approach to evaluating clinical efficacy, adverse events and drug Interactions of herbal remedies. *Phytother Res*. 2016;30:691–700.
- Bailey RL, Gahche JJ, Miller PE, Thomas PR, Dwyer JT. Why US adults use dietary supplements. *JAMA Intern Med*. 2013;173:355–361.
- US Government Accountability Office. Dietary supplements: FDA should take further actions to improve oversight and consumer understanding. <https://www.gao.gov/new.items/d09250.pdf>. Published 2009.
- US Senate. Dietary Supplement and Nonprescription Drug Consumer Protection Act. <https://www.congress.gov/109/plaws/publ462/PLAW-109publ462.pdf>. Published 2006.
- Knapik JJ, Trone DW, Austin KG, Steelman RA, Farina EK, Lieberman HR. Prevalence, adverse events, and factors associated with dietary supplement and nutritional supplement use by US Navy and Marine Corps personnel. *J Acad Nutr Diet*. 2016;116:1423–1442.
- Carvey CE, Farina EK, Lieberman HR. Confidence in the efficacy and safety of dietary supplements among United States active duty army personnel. *BMC Complement Altern Med*. 2012;12:182.
- Institute of Medicine. *Use of Dietary Supplements by Military Personnel*. Washington, DC: National Academy Press; 2008.
- Hoban CL, Byard RW, Musgrave IF. A comparison of patterns of spontaneous adverse drug reaction reporting with St. John's Wort and fluoxetine during the period 2000–2013. *Clin Exp Pharmacol Physiol*. 2015;42:747–751.
- Brown AC. Liver toxicity related to herbs and dietary supplements: online table of case reports. Part 2 of 5 series. *Food Chem Toxicol*. 2017;107:472–501
- Brown AC. Kidney toxicity related to herbs and dietary supplements: online table of case reports. Part 3 of 5 series. *Food Chem Toxicol*. 2017;107:502–519
- Brown AC. Heart toxicity related to herbs and dietary supplements: online table of case reports. Part 4 of 5. *J Diet Suppl*. 2018;15:516–555.
- Uppsala Monitoring Centre. *Safety monitoring of medicinal products: guidelines for setting up and running a pharmacovigilance centre*. Uppsala, Sweden: Uppsala Monitoring Centre; 2000.
- Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. *Clin Pharmacol Ther*. 2012;91:1010–1021.
- Honig PK. Advancing the science of pharmacovigilance. *Clin Pharmacol Ther*. 2013;93:474–475.
- Tatonetti NP, Fernald GH, Altman RB. A novel signal detection algorithm for identifying hidden drug–drug interactions in adverse event reports. *J Am Med Inform Assoc*. 2012;19:79–85.
- Avillach P, Dufour JC, Diallo G, et al. Design and validation of an automated method to detect known adverse drug reactions in MEDLINE: a contribution from the EU-ADR project. *J Am Med Inform Assoc*. 2013;20:446–452.
- Shetty KD, Dalal SR. Using information mining of the medical literature to improve drug safety. *J Am Med Inform Assoc*. 2011;18:668–674.
- Winnenburg R, Sorbello A, Ripple A, et al. Leveraging MEDLINE indexing for pharmacovigilance—Inherent limitations and mitigation strategies. *J Biomed Inform*. 2015;57:425–435.
- Haerian K, Varn D, Vaidya S, Ena L, Chase HS, Friedman C. Detection of pharmacovigilance-related adverse events using electronic health records and automated methods. *Clin Pharmacol Ther*. 2012;92:228–234.
- LePendou P, Iyer SV, Bauer-Mehren A, et al. Pharmacovigilance using clinical notes. *Clin Pharmacol Ther*. 2013;93:547–555.

22. Warrer P, Hansen EH, Juhl-Jensen L, Aagaard L. Using text-mining techniques in electronic patient records to identify ADRs from medicine use. *Br J Clin Pharmacol*. 2012;73:674-684.
23. Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling with word embedding cluster features. *J Am Med Inform Assoc*. 2015;22:671-681.
24. Sarker A, Ginn R, Nikfarjam A, et al. Utilizing social media data for pharmacovigilance: a review. *J Biomed Inform*. 2015;54:202-212.
25. Poulriot Y, Chiang AP, Butte AJ. Predicting adverse drug reactions using publicly available PubChem BioAssay data. *Clin Pharmacol Ther*. 2011;90:90-99.
26. Kuhn M, Letunic I, Jensen LJ, Bork P. The SIDER database of drugs and side effects. *Nucleic Acids Res*. 2016;44:D1075-D1079.
27. Liu M, Wu Y, Chen Y, et al. Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs. *J Am Med Inform Assoc*. 2012;19:e28-e35.
28. Food and Drug Administration. FDA Adverse Event Reporting System (FAERS). <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm>. Published 2016.
29. MedEffect Canada. Canada vigilance adverse reaction online database. <http://www.hc-sc.gc.ca/dhp-mps/medeff/databasdon/index-eng.php>. Published 2016.
30. Banda JM, Evans L, Vanguri RS, Tatonetti NP, Ryan PB, Shah NH. A curated and standardized adverse drug event resource to accelerate drug safety research. *Sci Data*. 2016;3:160026.
31. Xu H, Stenner SP, Doan S, Johnson KB, Waitman LR, Denny JC. MedEx: a medication information extraction system for clinical narratives. *J Am Med Inform Assoc*. 2010;17:19-24.
32. Wang L, Jiang G, Li D, Liu H. Standardizing adverse drug event reporting data. *J Biomed Semantics*. 2014;5:36.
33. UMLS. RxNorm. <https://www.nlm.nih.gov/research/umls/rxnorm/>. Published 2016.
34. Observational Health Data Science and Informatics (OHDSI). <http://www.ohdsi.org/>. Published 2016.
35. Sharma V, Sarkar IN. Identifying natural health product and dietary supplement information within adverse event reporting systems. *Pac Symp Biocomput*. 2018;23:268-279.
36. Sharma V, Restrepo MI, Sarkar IN. Solr-Plant: efficient extraction of plant names from text. *BMC Bioinformatics*. 2019;20:263.
37. Boyle B, Hopkins N, Lu Z, et al. The taxonomic name resolution service: an online tool for automated standardization of plant names. *BMC Bioinformatics*. 2013;14:16.
38. Gavalí DK, Kulkarni KS, Kumar A, Chakraborty BS. Therapeutic class-specific signal detection of bradycardia associated with propranolol hydrochloride. *Indian J Pharmacol*. 2009;41:162-166.
39. Giribet G. TNT: tree analysis using new technology. *Syst Biol*. 2005;54:176-178.
40. Hewett M, Oliver DE, Rubin DL, et al. PharmGKB: the pharmacogenetics knowledge base. *Nucleic Acids Res*. 2002;30:163-165.
41. Mizutani S, Noro Y, Kotera M, Goto S. Pharmacoepidemiological characterization of drug-induced adverse reaction clusters towards understanding of their mechanisms. *Comput Biol Chem*. 2014;50:50-59.
42. Sharma V, Sarkar IN. Bioinformatics opportunities for identification and study of medicinal plants. *Brief Bioinform*. 2013;14:238-250.
43. National Institute of Diabetes and Digestive and Kidney Diseases. *Livertox: Clinical and Research Information on Drug-induced Liver Injury*. Bethesda, MD: US National Library of Medicine; 2017.
44. Dietary Supplement Label Database (DSLDD). <https://dslid.nlm.nih.gov/dslid/index.jsp>. Accessed July 18, 2017.
45. Sarkar IN. Biodiversity informatics: organizing and linking information across the spectrum of life. *Brief Bioinform*. 2007;8:347-357.
46. Sharma V, Sarkar IN. Leveraging biodiversity knowledge for potential phytotherapeutic applications. *J Am Med Inform Assoc*. 2013;20:668-679.
47. Wassef B, Kohansieh M, Makaryus AN. Effects of energy drinks on the cardiovascular system. *World J Cardiol*. 2017;9:796-806.
48. Mougin F, Dupuch M, Grabar N. Improving the mapping between MedDRA and SNOMED CT. In: Peleg M, Lavrač N, Combi C, eds. *Artificial Intelligence in Medicine. AIME 2011. Lecture Notes in Computer Science*. Berlin, Germany: Springer; 2011:220-224. doi:10.1007/978-3-642-22218-4_27.
49. Cai M-C, Xu Q, Pan Y-J, et al. ADReCS: an ontology database for aiding standardization and hierarchical classification of adverse drug reaction terms. *Nucleic Acids Res*. 2015;43:D907-D913.
50. Planet PJ, DeSalle R, Siddall M, Bael T, Sarkar IN, Stanley SE. Systematic analysis of DNA microarray data: ordering and interpreting patterns of gene expression. *Genome Res*. 2001;11:1149-1155.
51. Tambaro S, Bortolato M. Interactions of Cannabis and Amphetamine-Type Stimulants. In: Campolongo P, Fattore L, eds. *Cannabinoid Modulation of Emotion, Memory, and Motivation*. New York, NY: Springer; 2015:409-442.