

Evaluation of Herbal and Dietary Supplement Resource Term Coverage

Nivedha Manohar^a, Terrance J. Adam^{a,b}, Serguei V. Pakhomov^{a,b}, Genevieve B. Melton^{a,c}, Rui Zhang^{a,c}

^a Institute for Health Informatics, University of Minnesota, MN, USA

^b College of Pharmacy, University of Minnesota, Minneapolis, MN, USA

^c Department of Surgery, University of Minnesota, Minneapolis, MN, USA

Abstract

The use of Complementary and Alternative Medicine (CAM) is increasingly popular in places like North America and Europe where western medicine is primarily practiced. People are consuming herbal and dietary supplements along with western medications simultaneously. Sometimes, supplements and drugs react with one another via antagonistic or potentiation actions of the drug or supplement resulting in an adverse event. Unfortunately, it is not easy to study drug-supplement interactions without a standard terminology to describe herbal and dietary supplements. This pilot study investigated coverage of supplement databases to one another as well as coverage by the Unified Medical Language System (UMLS) and RxNorm for supplement terms. We found that none of the supplement databases completely covers supplement terms. UMLS, MeSH, SNOMED CT, RxNorm and NDF-RT cover 54%, 40%, 32%, 22% and 14% of supplement concepts, respectively. NDF-RT provides some value for grouping supplements into drug classes. Enhancing our understanding of the gap between the traditional biomedical terminology systems and supplement terms could lead to the development of a comprehensive terminology resources for supplements, and other secondary uses such as better detection and extraction of drug-supplement interactions.

Keywords:

Herbal and nutritional supplements; Complementary and alternative medicine; Biomedical terminology; UMLS; RxNorm.

Introduction

The use of Complementary and Alternative Medicine (CAM) is rising in North America and Europe where conventional western medicine is practiced [1,2]. This is reflected by increasing sales of herbal supplements in the United States and Europe. For example, only 28 percent of women over 60 years in the United States took a Calcium supplement in the early 90s. A decade later, nearly 61 percent of women in that age group were taking Calcium supplements. In the United States, herbal supplement sales have reached an all-time high of 6 billion dollars per year according to a 2013 report published by the American Botanical Council [3]. However, the growing use of supplements is occurring with many adults simultaneously taking prescription medicines. Almost one in six adults in the US taking prescription medicines also takes an herbal supplement [4]. This poses a risk of unintentional interactions between the drug and supplement, especially since documentation is not consistent on supplements [1,3,4]. According to a report published by the Center for Disease Control (CDC), more than half of adults in the U.S take dietary supplements and the use of supplements is on the rise [3]. Fifty-six percent of Europeans are also found to have

consumed some form of CAM in 2012 [2]. CAM includes a variety of medical systems, therapies and mind-body interventions. For this paper, we have focused on the subcategory that includes dietary supplements called “biologically based therapies” [1].

Although herbal medicines have gained popularity, knowledge about herbal medicines has not gone up among physicians practicing conventional medicine [4]. In a survey conducted among 1,157 clinicians from around the world, about 75.5 % of them said that doctors are “poorly informed” about herbal medicines. Only 12.9 % said that they always check if their patient is taking herbal medicines before prescribing conventional drugs [5]. There have been some steps to reduce the gap between the use of herbal supplements by consumers and the knowledge of supplements from a clinical perspective.

While herbal supplements are widely believed to be a safe addition to conventional medicines, they can potentially react with western medicine. Herbal and other dietary supplements could increase or antagonize the actions of drugs or they could potentially change pharmacokinetics or pharmacodynamics of the drug or supplement itself [6]. For these reasons, detailed research of drug-supplement interactions is an important area.

In contrast to drug-supplement interactions, drug-drug interactions have been more widely studied and documented. One type of resource leveraged for this, are terminologies which provide a standard and well-defined manner of identifying and describing drugs. For example, RxNorm is a standardized system to describe clinical drugs, developed by the U.S. National Library of Medicine (NLM), which is also the preferred representation of drugs for health information exchange. The brand names, generic names and ingredients are listed along with the relationship between those entities [2,7]. In contrast, there is no well-accepted equivalent for herbal supplements making it more difficult to study drug-supplement interactions. This difficulty creates limitations for studies about drug-supplement interactions and is a substantial barrier for extracting information from the limited existing investigative literature. It is also possible to identify new drug-supplements interactions by extracting information from literature [8] and or collecting through a patient interview [9]. Most studies on drug-supplement interactions focus only on particular known combinations, like cranberry and warfarin, ginkgo and warfarin, or St. John’s wort and digoxin [6].

One foundational step towards identifying drug-supplement interactions is establishing a comprehensive terminology system for supplements. This purpose of this paper is to evaluate the supplement term coverage across various online supplements databases. Coverages of the Unified Medical Language System (UMLS) Metathesaurus, Medical Subject Headings (MeSH), Systematized Nomenclature of Medicine--Clinical Terms (SNOMED CT), RxNorm and National Drug

File - Reference Terminology (NDF-RT) for representing supplement terminology were also investigated.

UMLS and MetaMap

UMLS is a repository of biomedical vocabularies, which integrates over 2 million names and concepts. It also contains relationships between concepts and brings together many health and biomedical vocabularies and standards. The UMLS includes the NCBI taxonomy, Gene Ontology, MeSH, Online Mendelian Inheritance in Man (OMIM) and SNOMED CT among over 100 other vocabularies [10].

MetaMap is a Natural Language Processing (NLP) application developed by the U.S. NLM to map biomedical texts to the UMLS Metathesaurus concepts [11]. It is able to tag the text lexically and syntactically and identify the different concept candidates in the UMLS. In this study, MetaMap was implemented to map the supplement terms to UMLS concepts.

RxNorm and NDF-RT

RxNorm is a tool developed by the NLM. It has normalized names for branded and generic drug names. Drug names along with strength, brand names, dosage and ingredient information can be found in RxNorm. RxNorm also supports semantic interoperability between drug names and 15 other drug resources providing additional depth of information related to drugs absent with the UMLS [12].

NDF-RT is a part of the Veteran Health Administration's National Drug file. It is able to classify drugs into formal categories in addition to giving information about their molecular interactions, kinetics, therapeutic categories, and dose forms [13].

Supplement databases

Natural Standard Authority Database (NSAD) is a database set up by Natural Standard, an international research collaboration that reviews scientific literature on CAM. This database is arranged alphabetically and includes Spanish and French in addition to English and Latin names. Each entry includes supplement names in other languages, popular brand names, historical significance of the herb, indications for use, expert opinion on efficacy are all provided with citations from peer-reviewed journals [14].

Medscape is a website with medical content from the WebMD Health Professional Network. The Network strives to provide balanced, accurate health information. They provide information written by physicians and other authorities on particular topics making information from Medscape relatively reliable [15].

Natural Medicines Comprehensive Database (NMCD) is a database of nature-based medicines and alternative therapies managed by the therapeutic research center, composed of a panel of pharmacists and clinicians. The panel is not affiliated with any pharmaceutical company, and their work is intended to be useful for pharmacists and clinicians. The website is updated regularly and incorporates the latest research. The database has terms enabling term searches in several Asian and European languages, by botanical names and by common names. Under each plant substance, the effectiveness of the plant, safety, known interactions, mechanism of action, and adverse reactions are listed. There is an application with NMCD that allows users to search for interactions between a drug and a supplement. NMCD is aimed at professional clinicians or pharmacists and may not be easy to understand for a layperson [16].

MedlinePlus is the U.S. National Institute of Health's website for the general public. The website is produced by the NLM

and has information about diseases, conditions, health care issues drugs and supplements. The information in the database is available from the U.S. National Center for Complementary and Integrative Health (NCCIH). All statements are supported by evidence from peer-reviewed literature. Supplement names in languages other than English are not available for search. For this study, we only used information from the website on herbs and supplements [17].

Drugs.com is a website that aims to provide consumers with health, drugs and supplement-related information. Information is sourced from Harvard Health publications and Wolters Kluwer Health. Drugs.com provides indications, possible side effects, and interactions with other drugs. It is targeted at lay consumers and provided in a user-friendly format. The website also provides a platform for users to discuss with one another about the drugs. Having only 66 supplements, this was our shortest resource of supplements [18].

Methods

Comparison supplement terms coverages

We compared the supplement term coverage between all five databases in a several step process, as illustrated in Figure 1.

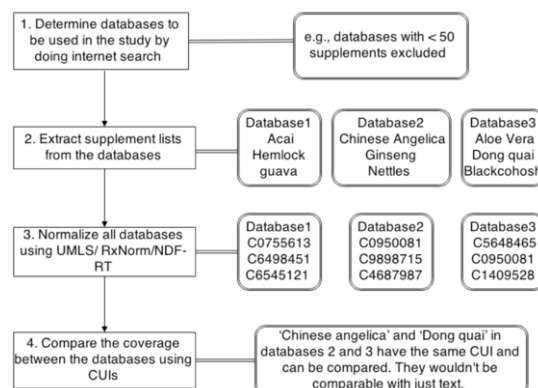


Figure 1 - Illustration of method to compare supplement terms across five online databases

Step 1: Determine the databases to be included in the study. An Internet search was performed for the top herbal supplement databases to identify the included databases. The top results were then subject to further exclusion criteria. If databases were not based on evidence-based science, they were rejected. For example, Holistic heart health, Healthline and WebMD's databases were rejected for not having peer-reviewed sources. Databases with very few (<50) supplements were also rejected. Also one other database on the Dr. Oz's website was no longer available, so that was excluded as well. Rxlist was also rejected since it merely mirrored NMCD's list. Finally, the supplements were obtained from each qualified resource website: NSAD, NMCD, MedlinePlus, Medscape, and drugs.com.

Step 2: Extract supplements from databases. We extracted supplement terms from online databases and used these in our further analysis. We manually removed redundant non-English terms from the list of NSAD while keeping their corresponding English names.

Step 3: Normalize supplements using UMLS, MeSH and SNOMED CT. To normalize supplement terms between different databases, supplements were first mapped to UMLS

Concept Unique Identifiers (CUIs) by using MetaMap – which then provides common meanings for the different sources of supplements. Only exact mapping with a score of 1000 were retained. Manual cleanup was performed to remove irrelevant mapping candidates. Semantic types such as ‘gene’ or ‘geographical area’ were eliminated before comparing the coverage between different databases. Irrelevant concepts after mapping were also excluded by manual review. When multiple CUIs were returned by MetaMap, the most accurate CUIs were retained. For example, when multiple CUIs were found for a single term, the CUI corresponding to semantic type “plants” was retained over “extracts”. Supplement terms were also mapped to MeSH and SNOMED by restricting the terminology source in MetaMap in an effort to use multiple sources to normalize the databases.

Step 4: Compare the coverage between the different databases. Before mapping, we first implemented exact match for supplement terms and analyzed how frequent an individual term exists in each database. UMLS CUIs were then used to compare the coverage of supplements between five databases. We further generate a Venn diagram showing the coverage across databases using UMLS. After integrating concepts from all databases into a full list, we also grouped UMLS concepts based on their semantic types.

Analysis of supplements representation in RxNorm and NDF-RT

To further investigate how RxNorm and NDF-RT represent supplements, we first used RxMix functions to map terms of each database to RxNorm and NDF-RT concepts. RxMix is an NLM developed tool to allow users to combine all functions of RxNorm, RxTerms, and NDF-RT [19]. We then integrated all mapped RxNorm concepts into a respective list. To further investigate the detailed drug class information, we used RxMix to analyze the integrated RxNorm concept list from all databases. Specifically, we created such a workflow to identify the VA drug class for a given supplements string:

RxNorm : findRxcuiBySting → RxNorm : getRelatedByType →

NDF – RT : findConceptsByID → NDF – RT : getVaClassOfConcept

Results

The first part of this study compared supplement terms without referring to any standard terminology source to investigate if any of the databases are able to cover all the supplements from the pooled lists of the databases. Figure 2 shows the frequency in which a particular supplement is found across the 5 databases. Without any term normalization, very few supplements are found across all five databases, but about 2,500 terms appeared in only one of databases.

To compare supplement term coverage across five databases, we normalized terms by using UMLS, MeSH, SNOMED CT, RxNorm, or NDF-RT source. The number of terms mapped to various sources from each database is shown in Table 1. In general, UMLS has the largest coverage for the supplements, and MeSH and SNOMED CT cover more supplement terms than RxNorm and NDF-RT. For example, the largest database NSAD had 1,331 total terms extracted from the website resulting in 1,088 UMLS concepts, 565 RxNorm concepts and 348 NDF-RT concepts with mapping. NMCD had a higher percentage coverage (78%) by UMLS concepts. MedlinePlus and Drugs.com were also covered well by all terminology sources. After integrating supplement terms from different databases into a comprehensive list, we obtained 3,115 unique terms, resulting in 1,683 UMLS, 1,257 MeSH, 984 SNOMED

CT, 677 RxNorm, and 422 NDF-RT concepts. Clearly, none of these terminologies can fully represent supplement terms.

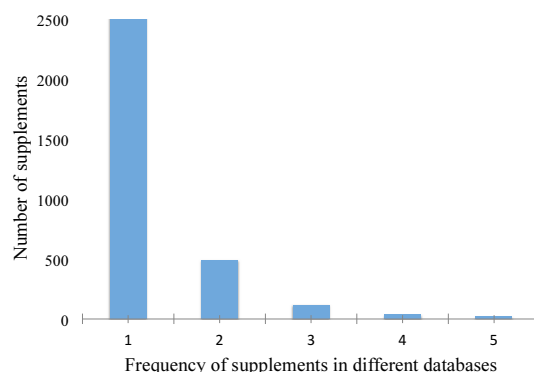


Figure 2 - Frequency of supplement terms across databases

Table 1 – Number of supplement terms mapped to the UMLS, MeSH, SNOMED CT, RxNorm and NDF-RT concepts

	NSAD	NMCD	MedScape	Medline	Drugs
Terms	1331	1038	1080	134	66
UMLS	1088	810	398	121	65
MeSH	930	658	307	108	49
SNOMED	669	492	291	99	41
RxNorm	565	400	228	93	60
NDF-RT	348	272	124	75	43

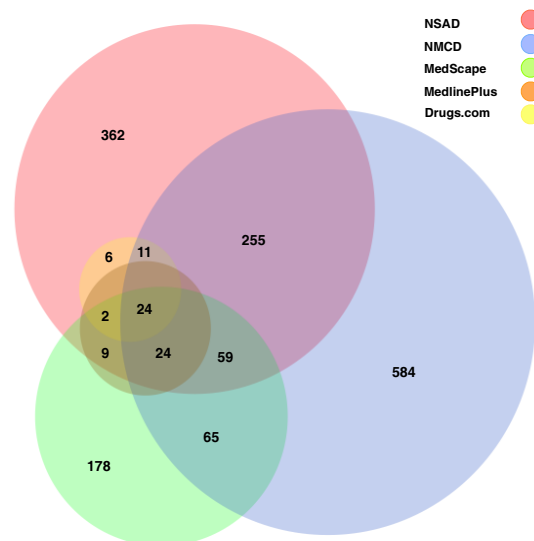


Figure 3 – Venn diagram of UMLS concepts for supplements across five databases

Figure 3 shows a Venn diagram of supplements concepts from the 5 databases after mapping to UMLS concepts. Twenty-four concepts are common to the 5 databases. All five databases have unique CUIs and different coverage with other databases. Not even the largest among the five databases is able to cover a large percentage of terms comprehensively.

After analyzing the semantic types in UMLS concepts, we found that 34% of concepts are plants (e.g., cat's claw, comfrey, damiana), and 18% are organic chemicals (Table 2). Many concepts belong to multiple semantic types, such as Aloe vera, which is a plant, a pharmacologic substance and an organic chemical. About 5% of concepts were classified as food.

Table 2 – Top semantic types of UMLS concepts for supplements

Semantic type	Percentage
Plant	34%
Pharmacologic Substance	27%
Organic Chemical	18%
Food	5%
Biologically Active Substance	4%
Element, Ion, or Isotope	2%
Vitamin	2%

With the aid of the tool RxMix, we found several drug classes for about 300 supplements (about 43% of mapped RxNorm concepts), some of which are listed in Table 3. Not surprisingly, multivitamins were the largest class for supplements. Some other classes, including stimulant laxatives, emollients, antihypoglycemics, antivertigo agents, genito-urinary agents, were also found. Since supplements in the same group can have similar properties, this help to investigate drug functions, adverse effects and interactions with other drugs.

Table 3 – VA drug Class for supplements in NDF-RT

Drug class	Examples
MULTIVITAMINS	alpha-tocopherol, biotin, coenzyme, cyanocobalamin, dexpantenol, ergocalciferol, folic acid, l-ascorbic acid, niacin, nicotinamide, pantothenol, phytonadione, pteroylglutamic acid, pyridoxine, retinol, riboflavin, thiamine
STIMULANT LAXATIVES	Aloe, cascara, cascara sagrada, castor oil, senna
RESPIRATORY AGENTS, OTHER	Camphor, eucalyptus oil, sodium chloride, turpentine oil
PANTOTHENIC ACID	Dexpantenol, pantothenic acid, pantothenol, Vitamin B5
EMOLLIENTS	aloe vera, glycerol, retinol, vitamin A, vitamin E
HERBS/ALTERNATIVE THERAPIES	docosahexaenoic acid, eicosapentaenoic acid, ginkgo, ginkgo biloba, inositol, l-methionine, methionine, vitamin E
GENITO-URINARY AGENTS, OTHER	dimethyl sulfoxide, povidone-iodine, yonon, yohimbe, yohimbine

Discussion

This study examines the problem of a lack of standard terminology in describing herbal and dietary supplements.

This is a significant problem because herbal supplements may interact with prescription medicines leading to potentially harmful interactions. Without a standardized terminology, studies and comparative research become more difficult, as well as other knowledge-sharing related to supplements.

Although many supplements and their interactions with drugs are published online, they usually use different names and focus on different entities. We found that all databases we examined had substantial gaps. The low degree of supplement terms commonality does create difficulty in utilizing a single database source for supplement study development and evaluation at present. A combined resource approach is likely needed for robust supplement term identification.

Many of the supplement terms could not be mapped to UMLS Metathesaurus. On review of some of the unmapped terms a number of findings were noted. Unlike traditional medications which typically have a generic and trade name and a chemical identifier, the dietary supplements are more complicated especially plant products. Starting with Latin Genus and Species information a particular supplement may have multiple common names depending on where the product is grown and the ultimate supplement consumer. Several common supplements had five or more common names making it challenging to map the name to a single entity. Several exact matches were also identified as irrelevant mappings. For example, cartilage (bovine and shark) was mapped to "1000 C0007301 Cartilage [Tissue]".

In addition, particular supplements may have different names based on qualifiers for the country of origin or the part of the plant which was used. Some products may have different reference names based on whether it was prepared by some means such as an extract which chemically is similar to the parent supplement but may be changed in some way after preparation. Arguably this may reflect more a dosage change in some cases but will vary on the supplement and the preparation method. Incorporation of the common preparation methods for supplement products may need additional exploration to identify the appropriate standards to accurately reflect the supplement being reviewed.

The number of concepts retrieved by the MetaMap application is much lower than the number of supplements in the NSAD and Medscape databases. One reason is that the MetaMap was also unable to map some chemical names (e.g., 1,3,7-trimethyl-2,6-dioxopurine). Another reason for the observed limited performance of MetaMap is that we restricted matches to exact mapping with a score of 1000. Some concepts with a lower score were therefore excluded. For example, "new zealand green-lipped mussel" was mapped to two UMLS concepts "933 NEW ZEALAND GREEN MUSSEL (New Zealand green mussel extract) [Organic Chemical, Pharmacologic Substance]" and "637 lipped (Lip structure) [Body Part, Organ, or Organ Component]". Another example is a combination of herbs that are available as a single supplement listing – for example, "bilberry/evening primrose/flax", which did not exactly map to any single concept. After manually reviewing those with lower mapping scores (<1000), we found additional 15, 20 and 31 correct mappings for NMCD, NSAD and Medscape, respectively. It was also found that UMLS doesn't have a trinomial classification of plants, making the CUIs less specific. For example, ginseng from different countries (e.g., Chinese ginseng and Korean ginseng) were mapped to the same CUI with outputs as "1000 C0949314: Chinese ginseng (*Panax ginseng*) [Plant]" and "1000 C0949314: Korean ginseng (*Panax ginseng*) [Plant]". Both "ginseng, American" and "ginseng, siberian" were mapped to "C0949314 ginseng (*Panax ginseng*)". In practice, ginseng from different geographical areas is used for different purposes and may have

potentially different active supplement concentrations – arguing for classifying these under different concepts. UMLS perhaps lacks the depth needed to represent herbal supplements. Also, some of the databases used in this study had repetitions containing the same herb in different languages. Place of supplement origin is unique for herbal substances. Such geographic origin information sometimes can be recognized by the supplements names. For example, ma huang has the botanical name *Ephedra sinica* where sinica stands for China. There is another species of Ephedra called *Ephedra funerea* named for the funeral mountains in the United States. Although UMLS had a broader coverage of supplement concepts, RxNorm and NDF-RT, as part of UMLS, can provide more detailed information about medications, including brand and generic name, dosage, drug class and interactions, which is useful for finding relationships between supplements. Future inclusion of additional supplements into these resources and frameworks will have significant value for secondary uses of supplement information. Our study has a number of limitations. One limitation is the use of only five databases, all of which are based in North America. A more comprehensive study would include more databases from geographically diverse sources. While the MetaMap system is accurate, the application is limited to English terms only, and only exact matches were included in this study, resulting in a large number of terms that did not map. The Anatomical Therapeutic Chemical (ATC) classification system [20] developed by the World Health Organization is another terminology we plan to evaluate the supplements term coverage. Future research will also focus on integrating knowledge resources to form a more comprehensive system of terminology for describing herbal and dietary supplements.

Conclusion

This pilot study compared supplements across five popular online herbal supplement databases. None of the databases were found to be comprehensive enough to represent a large majority overall of supplements in these resources. Analysis of the coverage of UMLS, MeSH, SNOMED CT, RxNorm and NDF-RT for describing herbal and dietary supplements demonstrated that UMLS covered over half (54%) of the supplements extracted and integrated from five databases. RxNorm and NDF-RT contain more detailed drug-related information, including drug classes for some supplements. This study is important because it provides insights on current gaps and potential opportunities to enhance our understanding of how different terminologies can be used to represent herbal supplements.

Acknowledgments

This research was supported by the University of Minnesota Informatics Institute on the Horizon grant (RZ), the Agency for Healthcare Research & Quality grant (#1R01HS022085-01) (GM), and University of Minnesota Clinical and Translational Science Award (#8UL1TR000114-02) (Blazer).

References

[1] Ventola CL. Current Issues Regarding Complementary and Alternative Medicine (CAM) in the United States. P&T 2010; 35(8): 461-8.

- [2] Zuzak TJ, Boňková J, Careddu D, Garami M, Hadjipanayis A, Jazbec J, et al. Use of complementary and alternative medicine by children in Europe: Published data and expert perspectives. Complement Ther Med 2013; 4:(21) Supplement 1(0): S34-47.
- [3] Gahche J, Bailey R, Burt V, Hughes J, Yetley E, Dwyer J, et al. Dietary Supplement Use Among U.S. Adults Has Increased Since NHANES III (1988–1994). April 2011; 2014 (Nov 4, 2014).
- [4] Kaufman D, Kelly J, Rosenberg L, Anderson T, and Mitchell A. Recent Patterns of Medication Use in the Ambulatory Adult Population of the United States. The Journal of the American Medical Association 2002; 287(3): 337-44.
- [5] http://dtb.bmj.com/site/about/DTB_survey_on_herbal_medicines.pdf.
- [6] Gardiner P, Phillips R, and Shaughnessy AF. Herbal and Dietary Supplement–Drug Interactions in Patients with Chronic Illnesses. American Family Physician 2008; 77(1): 73-8.
- [7] Bodenreider O, and Peters LB. A graph-based approach to auditing RxNorm. Journal of Biomedical Informatics 2009 April 2004; 42: 558-70.
- [8] Zhang R, Adam T, Simon G, Cairelli M, Rindflesch T, Serguei P, et al. Mining biomedical literature to explore interactions between cancer drugs and dietary supplements. AMIA 2015 Joint Summit on Translational science; Mar 2015; San Francisco.
- [9] Scarton LA, Zeng Q, Bray BE, and Shane-McWhorter L. Feasibility and potential benefit of collecting complementary and alternative medicine data through a computerized patient interview. AMIA Annual Symposium Proceedings, Washington, DC, 2011.
- [10] Bodenreider O. The unified medical language system (UMLS): Integrating biomedical terminology. Nucleic Acids Research 2004; 32:267-70.
- [11] Aronson AR, Lang F, Aronson AR, and Lang F. An overview of MetaMap: Historical perspective and recent advances. Journal of the American Medical Informatics Association 2010; 17(3): 229-36.
- [12] RxNorm. <http://www.nlm.nih.gov/research/umls/rxnorm/>
- [13] NDF-RT. <http://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NDFRT/>
- [14] Natural Standard Authority Database. <https://naturalmedicines.therapeuticresearch.com/>
- [15] MedScape. <http://www.medscape.com/public/about>
- [16] Natural Medicines Comprehensive Database. <http://naturaldatabase.therapeuticresearch.com/>
- [17] MedlinePlus. <http://www.nlm.nih.gov/medlineplus/druginformation.html>
- [18] Drugs.com. <http://www.drugs.com/support/about.html>
- [19] RxMix. <http://mor.nlm.nih.gov/RxMix/>
- [20] ATC. http://www.whooc.no/atc_ddd_index/

Address for correspondence

Rui Zhang, PhD, University of Minnesota, 420 Delaware Street SE, Minneapolis, MN, USA, 55455. Email: zhan1386@umn.edu