

# Comparing Existing Resources to Represent Dietary Supplements

Rubina F. Rizvi, MD, PhD<sup>1,2</sup>, Terrence J. Adam, MD, PhD, RPh<sup>1,2</sup>, Elizabeth A. Lindemann, BA<sup>5</sup>, Jake Vasilakes, MS<sup>1,2</sup>, Serguei VS. Pakhomov, PhD<sup>1,2</sup>, Jeffrey R. Bishop PharmD, MS, BCPP<sup>3,4</sup>, Genevieve B. Melton, MD, PhD<sup>1,5</sup>, Rui Zhang, PhD<sup>1,2</sup>

<sup>1</sup>Institute for Health Informatics, <sup>2</sup>Department of Pharmaceutical Care & Health Systems, <sup>3</sup>Department of Experimental and Clinical Pharmacology, <sup>4</sup>Department of Psychiatry, <sup>5</sup>Department of Surgery, University of Minnesota, Minneapolis, MN

## Abstract

*Dietary supplements, often considered as food, are widely consumed despite of limited knowledge around their safety/efficacy and any well-established regulatory policies, unlike their drug counterparts. Informatics methods may be useful in filling this knowledge gap, however, the lack of standardized representation of DS hinders this progress. In this pilot study, five electronic DS resources, i.e., NM, DSID & NHPID (ingredient level) and DSLD & LNHPD (product level), were evaluated and compared both quantitatively and qualitatively employing four phases. Essential data elements needed for comprehensive DS representation were compiled based on LanguaL code (food) & AHFSA (drugs) guidelines and employed as a check-list. We further investigated the completeness of DS representation by incorporating Ginseng and Fish oil as examples. We found fragmented and inconsistent distribution of DS representation in terms of essential data elements across five resources. This study provides a preliminary platform for development of standardized DS terminology/ontology model.*

## Introduction

Widespread use of Dietary Supplements (DS) is commonly seen among people belonging to different ethnicities, backgrounds, ages and genders across the globe despite insufficient evidence about their safety, efficacy and regulatory guidelines<sup>1,2</sup>. According to National Health and Nutrition Examination Survey data (NHANES) conducted between year cycles 1971–1974 and 2003–2006, the age adjusted consumption of DS has gradually increased in male (28% to 44%) and female (38% to 53%) consumer groups<sup>3</sup>. High consumption is especially seen among adults aged ≥60 y with 70% of older adults in the United States reported using one or more DS<sup>4</sup>.

Most people consider DS as safe and usually take them without consulting with healthcare providers. However, there is increasing evidence that DS could interact with prescription medicine and also cause serious adverse events. According to the Center for Disease Control and Prevention (CDC), an average of 23,000 annual emergency visits were related to DS in US<sup>5</sup>. Unlike prescription and over-the-counter medicines, DS are regulated by the Dietary Supplement Health and Education Act of 1994 (DSHEA)<sup>6</sup> and unlike drugs, clinical trials on DS safety and efficacy before getting marketed are not required. In addition, it is voluntary for healthcare providers to report only serious adverse events, including hospitalization, disability, and death, related to DS in the post-market surveillance. Current safety documentation for DS is very limited as most available information is based either on pharmacologic research, animal models or pharmaco-epidemiologic studies that often solely focus on a small set of drugs or supplements. These have greatly limited our ability to build up our knowledge on safety of DS.

Integrating information across diverse resources (e.g., online databases, biomedical literature) and further developing a data model to comprehensively represent DS and relevant safety information could potentially fill the knowledge gap to improve DS product safety. Common online sources for DS information include commercial databases such as Natural Medicines (NM), a primarily ingredient level resource, and publically available databases, such as U.S. Dietary Supplement Label Database (DSLDD) and Canadian Natural Health Product Ingredient database (NHPID), which are both considered as product level resources. Product labeling statements in these resources contain very limited safety information. There remains a critical need to more fully represent supplements by linking these databases using a common data model. Accurate and comprehensive supplement representation is also vital for accurate information extraction from both the biomedical literature and online databases.

Our prior work demonstrated number of gaps with regards to term representation in existing standard terminologies (Unified Medical Language System (UMLS), RxNorm and National Drug File-Reference Terminology (NDF-RT)) and clinical notes within Electronic Health Records <sup>7,8</sup>. Currently, no prior work has been done to comprehensively represent DS and the associated data model. In this study, we selected five online databases for DS, both at the ingredient and product level, and compared them against the list of data elements considered essential to comprehensively represent DS. This is the initial step towards the development of formal DS terminology/ontology model that could represent DS related information in a more accurate and consistent format, similar to drugs. The knowledge gained here will promote informatics research in DS, such as information extraction and knowledge discovery of safety on DS.

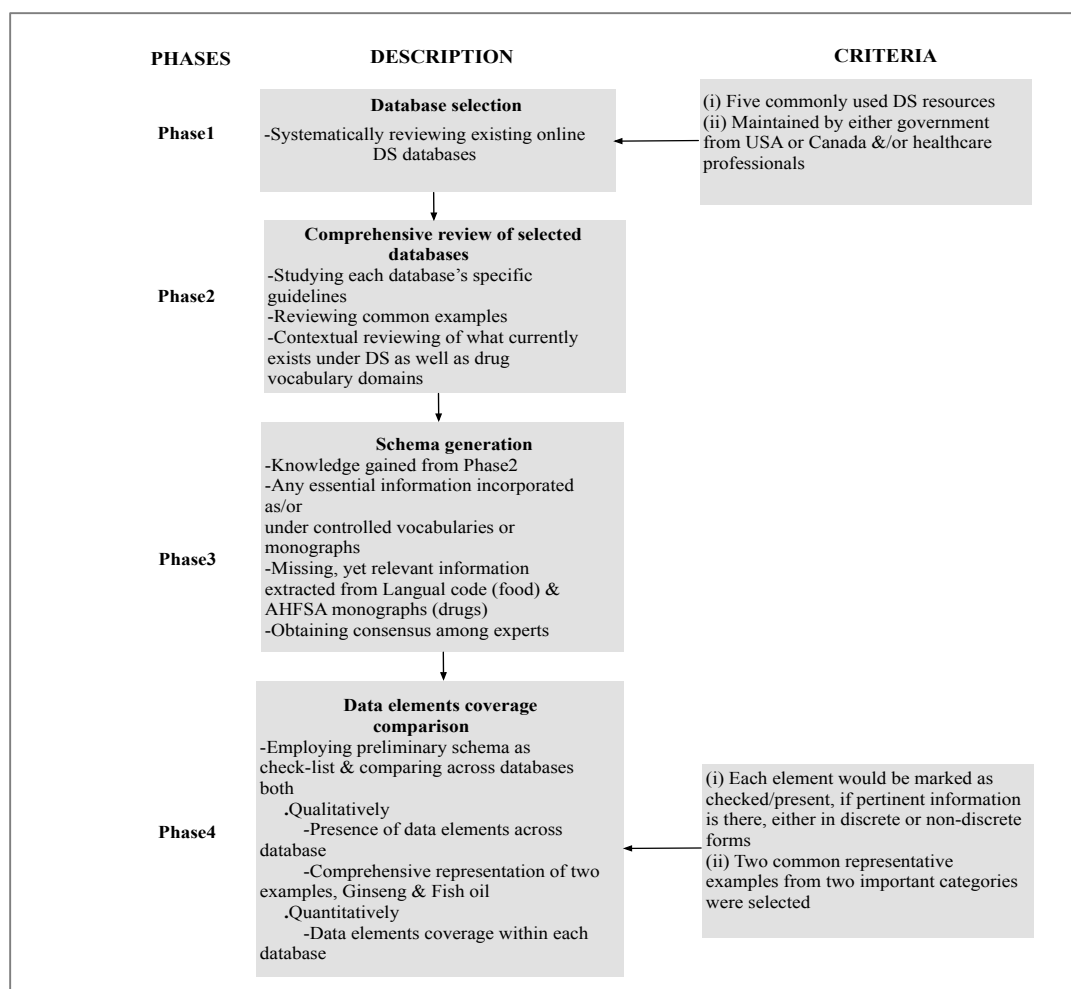
Current DS resources have term coverage at different levels of granularity (i.e., ingredient and/or product level), and are employed for various purposes by a wide range of users (e.g., pharmacist, physicians, manufacturers, etc.) The knowledge representation within these resources ranges from unstructured and fragmented to structured and comprehensive evidence based data having full monographs/controlled vocabularies and to the most structured and robust standardized terminologies. By definition, “monograph” is a document having detailed information on a concept while “controlled vocabularies” are the list of standardized terms employed for indexing and searching information for a particular concept while. Even within these terminological systems, variability exists in term as well as content coverage (related metadata coverage for each DS e.g., product name, active ingredient(s), drug strength and unit of measure, dosage form) <sup>9</sup>.

The objective of this study is to evaluate and compare existing online resources for DS representation. Common sources for DS information include databases such as, Dietary Supplement Ingredient Database (DSID) <sup>10</sup>, Natural Medicine (NM) <sup>11</sup>, Natural Health Products Ingredients Database (NHPID) <sup>12</sup>, Dietary Supplement Label Database (DSL) <sup>13</sup>, and Licensed Natural Health Products Database (LNHPD) <sup>14</sup>. Some drug databases also incorporate DS coverage to variable extent, e.g., DailyMed <sup>15</sup>, drugs.com <sup>16</sup>, etc., which are not included in this study. This goal is achieved through systematic review of selected resources at both the ingredient and product levels, compilation of the essential elements for DS representation as a preliminary model, and a comparison of the existing databases using this model as a check-list. To achieve this goal, we examined element of three ingredient level databases (DSID, NM, & NHPID) and two product level resources (DSL & LNHPD), to assess areas of uniqueness, overlaps and gaps where further information may be beneficial.

## **Materials and Methods**

### ***Study design***

The DS database comparison process was comprised of four phases: Phase1 - the selection of DS databases to be incorporated into our study; Phase2 - a top-down, systematic review of these databases to understand essential data elements required for DS representation; Phase3 - the generation of a preliminary model that could be employed to represent each DS concept in a consistent, precise, and holistic fashion; Phase4 - a comparison of data elements coverage across the selected databases. Figure 1 illustrates the phases and corresponding criteria.



**Figure 1.** Overview of the study design

### **Phase1 - Database selection**

We selected five electronic, evidence based DS resources at primarily two levels of representation: ingredient level (i.e., Dietary Supplement Ingredient Database (DSID), Natural Medicine (NM), Dietary Natural Health Products Ingredients Database (NHPID) and product level (i.e., Dietary Supplement Label Database (DSLID) and Licensed Natural Health Products Database (LNHPD)). Each of these databases were built exclusively for DS representation and created/maintained by either government from USA (DSID, DSLID) or Canada (NHPID, LNHPD), and/or healthcare professionals (NM-USA). More standardized drug terminological systems with some representation of DS were excluded from our preliminary study to be evaluated as the next step.

### **Phase2 - Review of databases**

A comprehensive, systematic review of the databases was performed by three co-authors: RR (health informaticist/physician), RZ (health informaticist), and TA (pharmacist/physician/health informaticist), by studying guidelines associated with each database and real time searching of a common DS as a representative example. The proposed LanguaL DS structured vocabulary (LanguaL™ DS)<sup>17</sup> thesaurus for DS was also reviewed in order to get additional contextual knowledge of what currently exists under drugs and DS vocabulary domain. LanguaL vocabulary is formulated by the US Federal Dietary Supplement Ingredient Database (DSID) *ad hoc* Working Group. In addition, a well-established and widely employed drug compendium, the American Hospital Formulary Service (AHSF) Drug Information (DI) Essentials<sup>18</sup>, was also reviewed. AHSF DI essentials is created by the American Society of Health-System Pharmacists (ASHP)<sup>19</sup>, an evidence based foundation for safe and effective drug therapy.

### **Phase3 – Data element model generation**

A comprehensive list of data elements to be used as a standard check-list was generated based on LanguaL code (food) & AHFSA (drugs) after an iterative process comprising an extensive review of DS and drug databases (both as controlled vocabularies, monographs or free text), their respective guidelines, and discussion among experts (RR, RZ, TA, JB). These data elements are defined in Table 1. It was found during this process that some data elements, while not explicitly present in a given database, are indeed expressed indirectly. For example, while none of the databases have a “lexicon variant” data element, alternate spellings for ingredients could be discovered by finding multiple synonymous entries that are spelled differently, e.g., “Ginkgo” and “Gingko”. Thus, a new facet, lexicon variant, could be created to explicitly express this concept. The model was further strengthened by including valuable inputs from experts, a team of informaticists, physicians and pharmacists (RZ, RR, TA, JB, GM, SP). The resulting data elements are all those deemed essential for comprehensive representation of DS.

**Table 1.** Data elements and their descriptions

<b>Data elements</b>	<b>Description</b>
<b>1. Title &amp; introductory information</b>	Basic description of type, sources, part of source, history, names and composition of the DS for the purpose of orientation and introduction
<b>2. Therapeutic intent (uses, indications, purpose of use, claims, effectiveness)</b>	Information on possible uses
<b>3. Mechanism of action</b>	Description of biochemical interaction resulting in pharmacological effect
<b>4. Pharmacokinetics</b>	Description on biotransformation and excretion inform of absorption, distribution, and elimination <sup>18</sup>
<b>5. Dosage &amp; administration</b>	Information on possible route of administration and dosage (strength/unit), adjusted for age, comorbidities, pregnancy/lactation
<b>6. Cautions</b>	Information on side effects, adverse reactions, contraindications, sensitivity reaction, interactions, toxicity/treatment, precautions for use under specific conditions (e.g., pregnancy, lactation, hepatic/renal impairment) or safety/interaction rating
<b>7. Packaging/ Manufacturing information</b>	Information on how DS was preserved, packaged (medium, contained), stability, amount and flavors. Any contact, copyright, tracking, labelling or licensing information is also included under this category
<b>8. Evidence based citations</b>	Any available references to well-conducted research

### **Phase4 - Data elements coverage comparison**

During this phase, the five selected databases were first qualitatively evaluated and compared for presences of essential data elements employing the above model as a check-list. Information pertaining to specific data elements was considered present if it existed in the controlled vocabularies (CV), monographs (M) or free text. We have also reported how DS should be represented comprehensively (by obtaining data from multiple resources) through two representative examples, Ginseng and Fish oil, cross checked with the model. A single variant of each example was employed for initial comparison.

For the quantitative evaluation, facets were first mapped to the data elements in Table 1 by a physician and health informaticist (RR). A coverage analysis was performed (JV) over the mapped facets, shown as the percentage of database entries having a non-null value. We use this evaluation to estimate the completeness of values for each data element for ingredients and products within each of these databases. Multiple facets in the source databases often map to a single data element in our schema. A given DS was considered to have data for a given ontology attribute if one or more of the corresponding facets in its source database was non-null.

## **Results**

Two levels of DS representation are given: (i) ingredient level (i.e., NM, DSID and NHPID) and (ii) product level (i.e., DSLD and LNHPD). The content under each database was assembled as either structured, controlled vocabulary, and/or descriptive monographs (Table 2).

**Table 2.** An overview of dietary supplement databases

Database	Developed by	Level of representation	Purpose	Content		Referenced standard
				*CV	**M	
<b>Dietary Supplement Ingredient Database (DSID)</b>	Nutrient data laboratory, USDA, NIH, ODS, NLM-USA	Ingredient level	Provides information on national estimates of ingredient content in various categories of DS.	x		Language code
<b>Natural Medicine (NM)</b>	Commercial database-USA	Ingredient level	Provides data related to various types of natural medicines, including food and herb & supplements.		x	Not available
<b>Natural Health Products Ingredient Database (NHPID)</b>	Health Canada-Canada	Ingredient level	Provide access to a scientific repository of approved natural health product ingredient information as well as monograph	x	x	TGA, ICH
<b>Dietary Supplement Label Database (DSLID)</b>	NIH, ODS, NLM-USA	Product level	Label information for over 55,000 DS currently marketed or off the market and consumed by National Health and Nutrition Examination participants in the latest survey	x		Language code
<b>Licensed Natural Health Products Database (LNHPD)</b>	Health Canada-Canada	Product level	A natural health products database provides access to a scientific repository of approved natural health product information as well as monograph.	x	x	TGA, ICH

**Note:** \*CV: controlled vocabulary; M: monograph; USDA: US Department of Agriculture; NIH: National Institute of Health; ODS: Office of Dietary Supplements; NLM: National Library of Medicine; TGA: Australian Therapeutic Goods Administration Approved Terminology for Medicines; ICH: International Conference on Harmonization.

Comparison of the five databases, employing the data element model as a check-list, showed a fragmented and inconsistent distribution of DS representation with variability in data elements distribution and as well as in the underlying content (Table 3), e.g., mechanism of action (only present in NM monographs), scientific name referring to Latin names (NM), or adverse reactions (NHPID, LNHPD) and adverse effects (NM) being used interchangeably.

**Table 3.** Evaluation of data element across existing dietary supplement databases

Sections	Data elements Facet	Databases (CV/M)				
		Ingredient level			Product level	
		DSID	NM*	NHPID*	DSLID	LNHPD*
		M*	CV	M*		
<b>1. Title &amp; introductory information</b>	1a. Type/Classification/Category (vitamin, mineral (or element), herb/botanical, amino acid/protein, other dietary substance supplementing the diet, metabolite, constituent, extract, isolate, or combination of any of these and combination of any of the above ingredients listed in 1–6 above)	✓	✓	✓	✓	✓

**Table 3.** Evaluation of data element across existing dietary supplement databases (continued)

Data elements		Databases (CV/M)				
Sections	Facet	Ingredient level			Product level	
		DSID	NM*	NHPID*	DSLID	LNHPD*
			M*	CV	M*	
	1b. Source (lower plant; animal; chemical; higher plant; bacteria; not identified)		✓	✓	✓	✓
	1c. Part of source (process-extract, concentrate etc.); anatomical part (animal, plant)		✓	✓	✓	
	1d. History/Country of origin		✓	✓		
	1e. Names: common name, variant spelling, proper names (scientific/Latin/chemical name), product name, synonyms	✓	✓	✓	✓	✓
	1f. Composition/Ingredients/Constituents (medicinal (single, combination), non-medicinal)	✓	✓	✓	✓	✓
<b>2. Purpose of use</b>	-		✓		✓	✓
<b>3. Mechanism of action</b>	-		✓			
<b>4. Pharmacokinetics</b>	Absorption/Distribution/ Elimination		✓			
<b>5. Dosage &amp; administration</b>	5a. Dose (general; adjusted for age/comorbidity/pregnancy/lactation)	✓	✓	✓	✓	✓
	5b. Administration (topical, oral, injectable etc.)	✓	✓	✓	✓	✓
	5c. Physical state/shape/form (tablet, capsule etc.)		✓	✓	✓	✓
	5d. Unit	✓	✓	✓	✓	✓
<b>6. Cautions</b>	6a. Side effects		✓		✓	✓
	6b. Adverse reactions		✓		✓	✓
	6c. Contraindications		✓		✓	✓
	6d. Sensitivity reaction		✓		✓	✓
	6e. Interactions		✓		✓	✓
	6f. Toxicity/Treatment		✓		✓	✓
	6g. Precautions for specific use (pregnancy, lactation, age dependent, hepatic/renal impairment)		✓		✓	✓
	6h. Safety/Interaction rating		✓			
<b>7. Packaging/ Manufacturing information</b>	7a. Formulation/Preservation (irradiation, gases etc.)				✓	✓
	7b. Packaging medium (alcohol, gas, other)					
	7c. Contact surface (glass, metal, paper etc.)					
	7d. Outside Package (bottle, blister packet etc.)				✓	
	7e. Stability (expiration, storage)				✓	
	7f. Net content				✓	
	7g. Proprietary flavors				✓	
	7h. Contact information				✓	
	7i. Brand Intellectual Property Statement				✓	
	7k. Labelling information				✓	
	7l. Regulations/Licensing information				✓	✓
<b>8. *EBM</b>			✓		✓	

**Note:** LNHPD\* has controlled vocabularies at primarily at product level full with a cross link to full monograph at the ingredient level. NHPID\* and NM\* have controlled vocabularies at ingredient level with a cross link to full monograph at the ingredient level. LNHPD and NHPID share same monographs. DSLID and LNHPD has additional information about product licensing and tracking.

Although the quantitative analysis at the level of sections showed inconsistent distribution of data elements across databases, there was high coverage (89%-100%) observed for each section within each database (Table 4). Since we did not normalize the ingredients, there exist some redundant instances.

**Table 4.** Completeness of selected data elements within existing databases

Sections	DSLID	LNHPD*	DSID	NM	NHPID*
<b>Title &amp; introductory information</b>	67456 (100%)	251757 (100%)	722366 (100%)	1136 (100%)	441407 (100%)
<b>Purpose of use</b>	66106 (98%)	223202 (89%)	-	1136 (100%)	-
<b>Mechanism of action</b>	-	-	-	1068 (94%)	-
<b>Dosage &amp; administration</b>	66106 (98%)	251757 (100%)	722366 (100%)	1045 (92%)	414923 (94%)
<b>Cautions</b>	62060 (92%)	251757 (100%)	-	1136 (100%)	-
<b>Packaging/ Manufacturing information</b>	67456 (100%)	251757 (100%)	-	-	-
<b>Evidence based medicine</b>	-	-	-	1136 (100%)	-

\* Data from monographs, e.g. pharmacokinetics information from NM, existed mainly as unstructured text, and would need further natural language processing to extract the information. Thus, it is not included in this study.

Two representative examples, Ginseng and Fish oil, were fit to the model in order to check the model's completeness and representative power. Data was obtained from various database in order to comprehensively represent the examples as shown in Table 5.

**Table 5.** Selected data elements to represent Ginseng and Fish oil using information in existing resources

Data elements		Representative example	
Sections	Facet	Ginseng	Fish oil
<b>1. Title &amp; introductory information</b>	1a. Type	Herb/botanical	Other dietary substance
	1b. Source	Plant (Araliaceae family)	mackerel, herring, tuna, halibut, salmon cod liver
	1c. Part of source (animal, plant)	Plant (Root/leaf/berry), an extract	Animal (Fish); an extract
	1d. History/Country of origin	Grows in Korea, northeastern China, and far-eastern Siberia. Pan am ginseng has been used medicinally in Asian countries, American Ginseng is a different herb	Goes back to Greenland Inuit people have a low incidence of heart disease despite a diet high in fat
	1e. Names	Ginseng (common name) Panax ginseng (scientific name) Chinese Red Ginseng, Korean Ginseng, Panax schinseng (synonyms) 21st Century Mega Multi For Men (product name)	Fish oil (common name) polyunsaturated long-chain fatty acids (scientific name) Menhaden Oil, Omega-3 Fatty acid (synonyms) Fish oil by 21 <sup>st</sup> century (product name)
	1f. Composition (medicinal, non-medicinal)	Panax ginseng (medicinal), could be single or combined Purified Water, Magnesium Stearate, Silicon Dioxide (non-medicinal)	Omega-3 Fatty acids (medicinal), could be single or combined Tocopherol, Gelatin, Glycerin, Soybean Oil (non-medicinal)
<b>2. Purpose of use</b>	-	Used as "adaptogen" for increasing resistance to environmental stress for treating Alzheimer's, Chronic Obstructive pulmonary disease, anemia	Hyperlipidemia, coronary heart disease, hypertension, bipolar disorder

**Table 5.** Selected data elements to represent Ginseng and Fish oil using information in existing resources (continued)

Data elements		Representative example	
Sections	Facet	Ginseng	Fish oil
<b>3. Mechanism of action</b>	-	Increases alcohol clearance in humans by 30% to 50%, likely by enhancing the metabolic activity of alcohol dehydrogenase	Effecting lipoproteins metabolism and also having antioxidant properties
<b>4. Pharmacokinetics</b>	Absorption/Distribution / Elimination	Absorbed into the blood over 24 hours, excreted in the urine only in trace amounts	Increases levels of omega-3 fatty acids in serum, plasma, and leukocyte, monocyte, myocardial, and erythrocyte phospholipids
<b>5. Dosage &amp; administration</b>	5a. Dose	2 tablets per servings Adults; Ginkgo leaf extract 120 mg for 4 weeks	Take 3 soft gels/day
	5b. Administration	Oral, topical and intravenous	Oral
	5c. Physical state/shape/form	Tablet	Soft gel capsule
	5d. Unit	mg	mg
<b>6. Cautions</b>	6a. Side effects	Insomnia, mastalgia, vaginal bleeding amenorrhea, hypertension, pruritus	Halitosis, heartburn, dyspepsia, nausea
	6c. Contraindications	Hemorrhagic or Thrombotic conditions	-
	6e. Interactions	Alcohol, anticoagulants, anti-diabetic drugs, estrogen, other herbs.	Anticoagulant, antihypertensive, contraceptives
	6f. Toxicity/Treatment	Lethal in high dosage	-
	6g. Precautions for specific use	Not safe among children, pregnant and lactating women and long-term use.	Like safe in pregnancy & lactation.
	6h. Safety/Interaction rating	Natural medicine safety rating “8”	Natural medicine safety rating “9”

## Discussion

Lack of a comprehensive DS knowledge representation has restricted research on their safety and efficacy. In our previous DS study, we identified mutual ingredients coverage for over the counter medicines and DS across several databases, despite them falling under different regulations and having different representations. In this study, we evaluated and compared five databases limited to DS only (both ingredient and product levels), by cross checking them against a preliminary, standardized set of data elements. This set of data elements entails information deemed essential to comprehensively represent DS. While, we found fragmented and inconsistent DS representations across these five databases (Table 3), high DS coverage (89%-100%) was discovered for each data element within a database (Table 4). The knowledge gained from this effort could be used as a platform to develop a more structured and comprehensive DS terminology/ontology system that could represent DS related information in a more accurate and consistent format, similar to their drug counter parts.

We observed varied levels of completeness in representation of DS across databases based on presence or absence of various data elements represented in the model. Several sections, such as title & introductory information, dosage, purpose of use, cautions, have wider coverage across databases as compared to other sections, such as mechanism of action, pharmacokinetics, and packaging. These discrepancies have multiple possible causes. For one, ingredient and product level resources are meant for different purposes and users. For example, product level databases provide more packaging related information. Another possible cause is the fragmented and/or incomplete coverage across databases,



e.g., NM is more comprehensive than DSID. Lexical variations were also observed in how data elements were referred to within a database, e.g., the terms “uses” and “purpose” employed interchangeably within NHPID. This variation was also observed across databases, e.g. use of term “adverse reactions” in NHPID/LNHPD vs. “adverse effects” in NM and use of word “purpose” in NHPID vs. “people use this for” in NM. Some overlap was seen among content coverage especially under naming sections. For example, the Latin names *Zingiber officinale* and *Panax ginseng* are (Latin terms) labelled as scientific names under NM. Whereas, the term synonym or taxonomical synonym universally represents alternate names. Content pertaining to some of the above facets is often inferred from other data rather than mentioned implicitly and is therefore marked as available, e.g., under DSLD and DSID, route of administration (oral) could be inferred from dose form (tablet).

Since DS usage is primarily self-initiated rather than based on clinicians’ recommendations, it renders unique challenges pertaining to efficacy, safety, and regulatory policies and practices to clinicians, consumers, and researchers<sup>1</sup>. In our set of data elements (Table 1), the section “cautions” includes following facets: side effects, adverse reactions, contraindications, sensitivity reaction, interactions, toxicity/treatment, precautions and safety/interaction ratings. Our DS evaluation showed substantial coverage for various facet under “caution” section across all five databases. Comprehensive information on DS risks could be made available through our DS ontology after cross mapping each facet under “caution” section to the pertinent databases.

Quantitative analysis (Table 4) showed fragmentary coverage of data elements across five databases, NM having most comprehensive coverage (7/8 sections), while DSID the least (2/8 sections). High percentage of data elements coverage (89%-100%) was observed within each database, thus making it possible to collect DS representation information from these databases. We found some unique data elements usually not seen in drug databases, such as source (animal or plant) and origin (Asia or North America), which can affect the functions of DS, e.g., American Ginseng vs. Korean Ginseng.

Complete representation of DS as shown for Ginseng and Fish oil, employing consistent vocabulary (Table 5 showing only important data elements at ingredient level) across databases, could help in more effective and efficient use of DS databases, e.g., mapping, integrating and searching for information.

There are few limitations associated with this study. Only five databases limited to USA and Canada were evaluated. Each section in the model was not weighted and that could potentially influence interpretation of results. It is also possible that we missed some data elements, not presented in these five databases. More granular analyses are needed for data sections as name categories, dose administration routes, dose forms etc. This research study is an initial step towards the development of more formal DS terminology/ontology model that could represent DS related information in more accurate and consistent format, similar to their drug counter parts. The knowledge gained here will promote informatics research in DS, such as information extraction and knowledge discovery of safety on DS. Other information, such as safety information, is largely stored in unstructured format in various data sources, which would prove difficulties to support clinical research in dietary supplement. Further efforts, such as extraction of structured information in standardized format from texts in resources, are required.

## Conclusion

This pilot study evaluated and compared DS data elements coverage by five DS databases, both from the USA (NM, DSLD and DSID) and Canada (NHPID and LNHPD). A model based on LanguaL code and AHFS guidelines was generated and employed as a checklist. Across all five databases, we found fragmented and inconsistent distribution of DS representation with variability in data elements and as well as in the underlying content. This study unveils the gaps in existing DS supplement domain and provides us with an essential knowledge to be used as a platform to build more standardized DS terminology/ontology model.

## Acknowledgements

This research was supported by National Center for Complementary & Integrative Health Award (#R01AT009457) (Zhang), the Agency for Healthcare Research & Quality grant (#1R01HS022085) (Melton), the National Center for Advancing Translational Science (#U01TR002062) (Liu/Pakhomov/Jiang), and the University of Minnesota Clinical and Translational Science Award (#8UL1TR000114) (Blazer).

## References

1. Institute of Medicine . Committee on the Use of Complementary and Alternative Medicine by the American P. Complementary and alternative medicine in the United States. National Academies P, editor. Washington, D.C: National Academies Press; 2005.
2. Bent S. Herbal medicine in the United States: review of efficacy, safety, and regulation: grand rounds at University of California, San Francisco Medical Center. *J Gen Intern Med.* 2008;23(6):854-9.
3. Bailey RL, Gahche JJ, Lentino CV, Dwyer JT, Engel JS, Thomas PR, et al. Dietary supplement use in the United States, 2003-2006. *J Nutr.* 2011;141(2):261-6.
4. Gahche JJ, Bailey RL, Potischman N, Dwyer JT. Dietary Supplement Use Was Very High among Older Adults in the United States in 2011-2014. *J Nutr.* 2017; 148 (10): 1968-1976
5. Geller AI, Shehab N, Weidle NJ, Lovegrove MC, Wolpert BJ, Timbo BB, et al. Emergency Department Visits for Adverse Events Related to Dietary Supplements. *N Engl J Med.* 2015;373(16):1531-40.
6. Dietary supplement health and education act of 1994 [Available from: [https://ods.od.nih.gov/About/DSHEA\\_Wording.aspx](https://ods.od.nih.gov/About/DSHEA_Wording.aspx) .
7. Manohar N, Adam TJ, Pakhomov SV, Melton GB, Zhang R. Evaluation of Herbal and Dietary Supplement Resource Term Coverage. *Stud Health Technol Inform.* 2015;216:785-9.
8. Zhang R, Manohar N, Arsoniadis E, Wang Y, Adam TJ, Pakhomov SV, et al. Evaluating Term Coverage of Herbal and Dietary Supplements in Electronic Health Records. *AMIA Annu Symp Proc.* 2015;2015:1361-70.
9. Saitwal H, Qing D, Jones S, Bernstam EV, Chute CG, Johnson TR. Cross-terminology mapping challenges: a demonstration using medication terminological systems. *J Biomed Inform.* 2012;45(4):613-25.
10. Dietary Supplement Ingredient Database (DSID) [Available from: <https://dietarysupplementdatabase.usda.nih.gov/>.
11. Natural Medicine (NM) [Available from: <https://naturalmedicines.therapeuticresearch.com/>.
12. Natural Health Products Ingredients Database ( NHPID ) [Available from: <http://webprod.hc-sc.gc.ca/nhp/ndb/dipsn/search-rechercheReq.do>.
13. Dietary Supplement Label Database (DSLID) [Available from: <https://www.dsld.nlm.nih.gov/dsld/index.jsp>.
14. Licensed Natural Health Products Database (LNHPD) [Available from: <https://www.canada.ca/en/health-canada/services/drugs-health-products/natural-non-prescription/applications-submissions/product-licensing/licensed-natural-health-products-database.html>.
15. DailyMed [Available from: <https://dailymed.nlm.nih.gov/dailymed/>.
16. Drugs.com [Available from: Drugs.com. <http://www.drugs.com/support/about.html>.
17. Saldanha LG, Dwyer JT, Holden JM, Ireland JD, Andrews KW, Bailey RL, et al. A structured vocabulary for indexing dietary supplements in databases in the United States. *J Food Compost Anal.* 2012;25(2):226-33.
18. American Hospital Formulary Service (AHSF) DI Essentials [Available from: <http://www.ahfsdruginformation.com/ahfs-di-essentials/>.
19. American Society of Health-System Pharmacists [Available from: <https://www.ashp.org/>.