# Identifying natural health product and dietary supplement information within adverse event reporting systems

Vivekanand Sharma and Indra Neil Sarkar
*Center for Biomedical Informatics, Brown University*
*Providence, RI 02912, USA*
*Email: vivekanand_sharma@brown.edu and neil_sarkar@brown.edu*

Data on safety and efficacy issues associated with natural health products and dietary supplements (NHP&S) remains largely cloistered within domain specific databases or embedded within general biomedical data sources. A major challenge in leveraging analytic approaches on such data is due to the inefficient ability to retrieve relevant data, which includes a general lack of interoperability among related sources. This study developed a thesaurus of NHP&S ingredient terms that can be used by existing biomedical natural language processing (NLP) tools for extracting information of interest. This process was evaluated relative to intervention name strings sampled from the United States Food and Drug Administration Adverse Event Reporting System (FAERS). A use case was used to demonstrate the potential to utilize FAERS for monitoring NHP&S adverse events. The results from this study provide insights on approaches for identifying additional knowledge from extant repositories of knowledge, and potentially as information that can be included into larger curation efforts.

*Keywords:* Natural Health Products; Dietary and Herbal Supplements; Adverse Event Detection; Terminology Mapping; Natural Language Processing.

## 1. Introduction

The biomedical community has benefitted from continuous development and improvement of automated methods for knowledge acquisition from heterogeneous data sources. A fundamental requirement for such tasks includes identification of entities of interest and their resolution to standard terminologies[1]. The process of converting unstructured free text fields from data into structured format creates opportunities to attain actionable knowledge by designing analytic enquiries. The heterogeneity of data from different sources poses challenges when seeking to perform comprehensive, multi-source analyses. Previous studies have demonstrated the utility of interlinked data from multiple sources to identify potential new knowledge[2–4]. The growing amounts of biomedical data from multiple sources suggest that an essential prerequisite for biomedical knowledge discovery will be the potential to leverage terminology resources for facilitating efficient indexing and subsequent retrieval. The biomedical domain is equipped with standard vocabularies from several sources that are used to facilitate access, retrieval and analysis of data from disparate data and knowledge sources. For example, the Unified Medical Language System (UMLS) Metathesaurus, maintained by the National Library of Medicine (NLM), is a repository of over one million biomedical concepts from more than 100 sources[5].

To support standardization and integration of available information about drugs and health related outcomes, the Observational Health Data Sciences and Informatics (OHDSI) workgroup was established with the goal of developing an open-source standardized knowledge base[6]. The

most significant utility of such a knowledge base is its ability to facilitate rigorous and accurate assessment of relationships between drugs and Health Outcomes of Interest (HOI). The Adverse Event Open Learning through Universal Standardization (AEOLUS) is a major product of the OHDSI workgroup, designed as a resource for the biomedical community[7]. AEOLUS consists of a standardized representation of FAERS[8] data, including normalization of drug names and health outcomes from the adverse event reports and precomputed common statistics. Use of FAERS data requires pre-processing, cleaning, and standardization, which presents a challenge for researchers intending to attain insights from adverse event reports. AEOLUS directly addresses this challenge, and reduces the requisite time and effort required to pursue research that utilize FAERS data.

Given the resources and initiatives in the biomedical and observational data realms to support a range of analyses, investigation of drug-HOI signals shows tremendous promise. However, the limited potential to investigate similar efficacy and safety issues related to dietary and herbal supplements (DHS) is generally due to lack of such resources. Although generally considered safe, there is evidence of DHS causing physical and economic harm[9]. An estimate of DHS-related adverse events suggest that they are associated with approximately 23,000 emergency department visits per year[10]. The incidents are higher for groups where the use of DHS is prevalent (e.g., among Navy and Marine Corps personnel, 22% of DHS users reported one or more adverse effects[11]). Such statistics underscore the need for systematic studies and the evaluation of available documentation in literature or reports associated with DHS use. The utility of existing biomedical vocabularies has been evaluated in the context of DHS, showing that UMLS generally, and, more specifically MeSH, SNOMED-CT, RxNorm, and NDF-RT include only 54%, 40%, 32%, 22%, and 14% of supplement concepts respectively[12]. Lack of robust acquisition of supplement documentation from electronic health records resulting from the gap between supplement and standard terminologies has also been highlighted[13]. Wang *et al.* found that only 14.67%, 19.65%, and 12.88% of ingredient terms from the Dietary Supplement Label Database (DSLD)[14] were mapped by UMLS, RxNorm, and NDF-RT, respectively[15]. The issue of less than 100% drug name mapping coverage in AEOLUS is noted by Banda *et al.* and attributed to those not found in RxNorm which include non-prescription products among other reasons resulting in unmapped records[7].

This study examined the potential and utility of creating a list of terms and concepts from ten sources that provide coverage for ingredients from Natural Health Products (NHP) and DHS. Using this as a resource, a custom thesaurus was built and used by the MetaMap NLP tool[16] to identify name strings (e.g., prescription drug, recreational substance, natural products and dietary supplements) found in FAERS. The system was specifically evaluated for its ability to recognize mentions of Natural Health Products and Supplements (NHP&S). The results from this study reveal challenges and opportunities in the development of an NHP&S terminology resource for automating acquisition of relevant information. The insights gained from this study may serve as motivation for improvement of the NHP&S thesaurus as well as its use for identifying and mapping relevant information in knowledge sources.

## 2. Materials and Methods

The goal of this study was to build a thesaurus of terms indicating NHP and DHS ingredients that could be integrated with extant biomedical Natural Language Processing (NLP) tools to facilitate acquisition of domain relevant information. Terms were identified from ten sources that included biomedical terminology sources as well as sources aimed at providing NHP&S information for healthcare providers and the general public. The terms were organized into concepts and a custom thesaurus that was subsequently used by the MetaMap[16] NLP tool to identify NHP&S concepts in FAERS. The utility of the NHP&S thesaurus was evaluated on randomly sampled intervention name strings. The NHP&S thesaurus was then used to process intervention name strings from FAERS to identify reports of adverse events associated with ingredients from NHP&S. A general overview of the approach is graphically depicted in Figure 1.

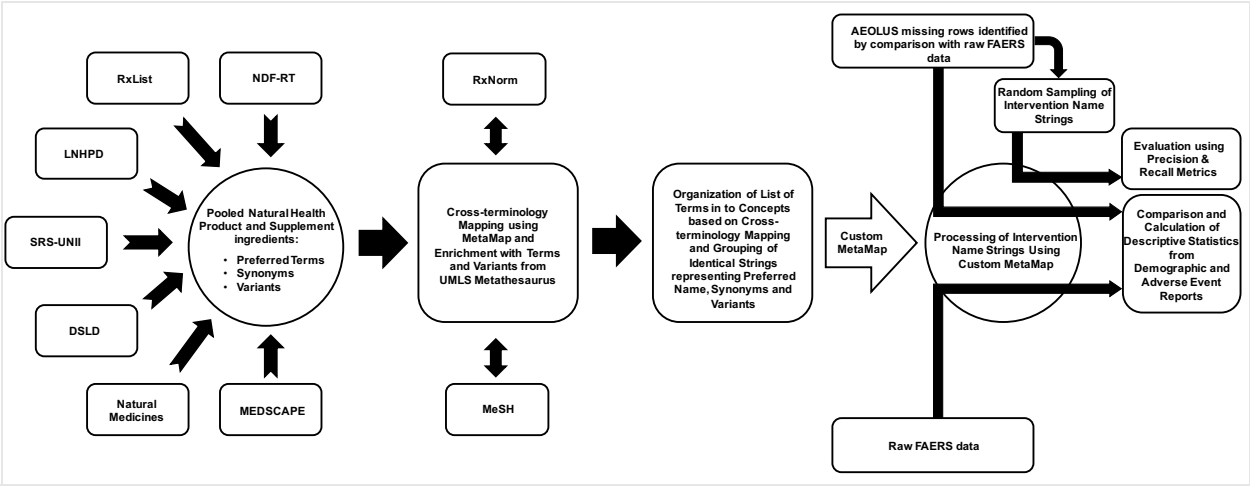


Figure 1. Overview of approach followed in this study.

### 2.1. *Identification of sources and extraction of NHP&S terms*

The goal of this step was to incorporate NHP&S terms from sources that offer reliable and comprehensive coverage of relevant terms, synonyms, and variants. The supplement ingredient terms and their synonyms were extracted from the databases shown in Table 1 (except RxNorm and MeSH, that was included after mapping). When available, the term variants were also extracted. Source identifiers were retained throughout the process; pseudo-identifiers were assigned for those terms that did not have an explicit identifier.

### 2.2. *Cross-terminology mapping of term strings from different sources*

The list of ingredient terms, synonyms, and variants from different sources were processed using the MetaMap NLP tool, which maps term strings to concepts included within UMLS Metathesaurus. From the MetaMap output (MMO), the mapped concepts, scores, semantic types, and terminology source(s) were extracted. The resulting list was filtered to retain those concepts that were identified with a perfect score of 1000. The filtering step included retaining concepts of semantic types associated with supplement ingredients[12]: (1) Plant (*plnt*); (2) Pharmacologic

Substance (*phsu*); (3) Organic Chemical (*orch*); (4) Food (*food*); (5) Biologically Active Substance (*bacs*); (6) Element, Ion, or Isotope (*elii*); and (7) Vitamin (*vita*). If the terminology source list included NDF-RT, RxNorm, or MeSH, it was recorded along with source identifiers.

Table 1. Sources selected for compiling NHP&S ingredient term list.

| Source | Description |
|---|---|
| LNHPD[17] | *Licensed Natural Health Product Database*: Contains information about NHPs that have been issued a product licence after quality, safety and efficacy assessment by Health Canada. This database provides medicinal and non-medicinal ingredient information for a variety of NHPs. |
| DSLD[14] | *Dietary Supplement Label Database*: This database of full label information is a result of collaboration between the Office of Dietary Supplement (ODS) and the NLM to serve as an educational and research tool for students, healthcare providers, and the public. |
| SRS-UNII[18] | *Substance Registration System - Unique Ingredient Identifier*: This resource provides unique ingredient identifiers for substances in drugs, biologics, foods, and devices. From among the list of ingredients, those that had taxonomic links (other than viruses) were retained. |
| RxList[19] | *RxList* is a resource that offers pharmacological information on drugs and supplements. As a part of the WebMD network, the content is updated with recent articles and data from reliable sources such as pharmacists, physicians, FDA, etc. The "supplements" section was used to gather listed terms. |
| Natural Medicines[20] | *Natural Medicines*: This resource combines features and functionality from two of the major natural medicine databases, Natural Standard and Natural Medicines Comprehensive Database. The section "Food, Herbs & Supplement" was used to gather terms of interest. |
| Medscape[21] | *Medscape*: In addition to prescription drugs, this resource contains information related to herbals and supplements categorized by therapeutic classes are also provided. The section "Herbals and Supplements" was used to gather study relevant terms. |
| NDF-RT[22] | *National Drug File - Reference Terminology*: This resource is a formal representation of a drug list that include ingredients and provides hierarchical drug classification. The categories include "Herbs/Alternative Therapies" which was relevant for this study. |
| RxNorm[23] | *RxNorm* provides normalized names for drugs and its links to several other drug vocabularies used in pharmacy management. In addition to prescription drugs, it also includes food and dietary supplements among other types of interventions. |
| MeSH[24] | *Medical Subject Headings*: This is a controlled vocabulary maintained by the NLM for indexing biomedical artifacts (e.g., biomedical literature). It includes a range of terms including those used for drugs and herbs. |
| UMLS[5] | *Unified Medical Language System* maintained by the NLM provides a unified repository for over one million inter-related biomedical concepts from more than 100 sources. MetaMap was used to map the term strings from sources listed above except RxNorm and MeSH. For the identified UMLS concept list all synonymous terms and variants were extracted. Mappings to RxNorm, MeSH and NDF-RT were included in the final thesaurus to equip this resource with the ability to provide references to those sources when processing text. |

## 2.3. *Grouping term strings and custom thesaurus*

Following MetaMap processing and identification of similar strings from across different sources, the list was enriched by extracting preferred terms, synonyms and variants from RxNorm, MeSH, as well as the UMLS Metathesaurus more generally (RxNorm: 4386 strings; MeSH:2826 strings; and Other sources: 5439 strings). Term strings indicating same ingredient across multiple sources

were grouped together. In addition to the entries mapped by MetaMap, the final dictionary included the unmapped terms which constitutes the major portion of the entries. Unique identifiers were assigned to indicate unique strings, variants, and concepts. This list was filtered against prescription drug list from National Drug Code Directory (NDC)[25] to remove any such ingredient name strings. This dataset was organized into tables for use with MetaMap Data File Builder suite (2016) to create a custom thesaurus that could be used with MetaMap ("Custom MetaMap").

### 2.4. *Sampling and evaluation*

The evaluation assessed the ability to recognize NHP&S strings as well as correctly eliminating non-NHP&S strings. A pool of unique strings was generated from missing rows of AEOLUS (mAEOLUS) by comparison with raw FAERS data files for years 2004-14. For example, ISR number 7811738 is missing in AEOLUS which contains NHP&S terms such as Red yeast rice, Fish oil, and Vitamin B6. Statistically significant random samples (95% Confidence Level at 4% Margin of Error) were selected and manually annotated as '*NHP&S*' or '*Non-NHP&S*'. To be more inclusive of NHP&S containing strings and to make evaluation more robust we performed the evaluation twice by creating two types of sample sets: (1) The strings were grouped separately according to year and random samples were picked without replacement; and (2) From the entire set of unique strings from the dataset, random samples were selected without replacement in ten iterations. The sampled strings were processed using Custom MetaMap and mappings with a score of 1000 (for MetaMap algorithm and scoring criteria refer to article by Aronson[26]) were retained to provide a more stringent evaluation and comparative performance in terms of our pipeline's ability to extract NHP&S records from FAERS when compared to those that were missed in AEOLUS as a result of inadequate coverage. True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) was scored based on the ability to distinguish between NHP&S and non-NHP&S related terms and on whether a given intervention name string mapped to NHP&S term (correctly or incorrectly). The annotation was performed by an annotator whose expertise is in the area of natural health products and dietary supplements (VS) under the supervision of another subject-matter expert in biomedicine (INS). Evaluation was assessed based on the standard metrics of Precision and Recall. Year-wise evaluation was performed on the first set of samples and mean and standard deviation was calculated. For the second type of sample, Precision and Recall were calculated on the pooled data from ten iterations.

### 2.5. *Identification and summarization of Adverse Drug Event (ADE) reports*

Intervention name strings from AEOLUS missing rows (mAEOLUS) and raw FAERS data were processed separately using Custom MetaMap. The use of FAERS rows that were missing from AEOLUS was used to test the hypothesis that, due to inadequate coverage of existing biomedical terminologies, there is loss of information related to NHP&S. The NHP&S annotated strings were mapped back to ADE reports and a comparative examination was done. A basic comparison of NHP&S and non-NHP&S reports was performed by isolating relevant demographic and adverse event data. The stratified counts were normalized with the total counts for a given group (NHP&S or non-NHP&S) and used for comparison. The System Organ Class (SOCs) associated with

Preferred Terms for ADEs (which are encoded using MedDRA[27]) were identified and a comparative summary at the level of SOC was calculated using normalized counts.

## 3. Result

### 3.1. *Identification of sources and extraction of NHP&S terms*

Compilation of strings was initially from seven sources: LNHPD, DSLD, SRS-UNII, RxList, Natural Medicines, Medscape, and NDF-RT. Cross-terminology mapping resulted in enrichment of terms from RxNorm, MeSH, as well as UMLS. The final groupings resulted in 81,680 concepts encompassing 320,579 strings. The counts shown in Table 2 are based on the sum of entry terms, scientific names, synonyms, vernaculars as well as variants. The Source IDs indicate the preferred term used to list a given NHP&S within a given source; String IDs are additional synonyms or variants. The total of Source ID counts for individual source was more than that in the final thesaurus due to overlapping terms from different sources. The counts of overlapping terms among the sources in listed in Supplemental Table 2.

Table 2. Counts of ingredient name strings extracted from included sources.

| Source | | Source IDs | String IDs |
|---|---|---|---|
| LNHPD | | 6,108 | 9,359 |
| DSLD | | 43,093 | 43,093 |
| SRS-UNII | | 15,492 | 165,786 |
| RxList | | 11,967 | 33,783 |
| Natural Medicines | | 1,208 | 1,208 |
| Medscape | | 193 | 1,248 |
| UMLS | NDF-RT | 4,179 | 11,273 |
| | RxNorm | 4,386 | 32,111 |
| | MeSH | 2,826 | 20,230 |
| | Other | 5,439 | 17,901 |

### 3.2. *Evaluation of Custom MetaMap*

The average number of distinct intervention name strings organized by year (2004-14) was 53,717.91±6796.87, ranging between 43,827 and 64,301. The sample size for random sampling from each year ranged between 592 and 595, with a mean of 593.45±0.89. The processing of sampled strings with Custom MetaMap resulted in mean precision and recall values of 0.94±0.01 and 0.72±0.08, respectively (F-score: 0.81±0.05). The total number of TP, FP, TN, and FN were 606, 39, 5640, and 241 respectively. Statistics for each individual year is provided in Supplemental Table 1. The second set was selected from 342,859 distinct intervention name strings from all years. The randomly selected sample size was 5990 gathered in ten iterations. Custom MetaMap processing of this sample resulted in a precision and recall of 0.93 and 0.66, respectively (F-score: 0.77). The total number of TP, FP, TN, and FN from this sample were 557, 40, 5102, and 291 respectively. A summary of the evaluation scores is listed in Table 3.

Table 3. Evaluation of Custom MetaMap on sampled intervention name strings from FAERS.

| | | Precision | Recall | F-score |
|---|---|---|---|---|
| Sampled year-wise | Range | 0.92-0.97 | 0.57-0.85 | 0.71-0.90 |
| | Mean | 0.94±0.01 | 0.72±0.08 | 0.81±0.05 |
| Sampled in 10 iterations then pooled | | 0.93 | 0.66 | 0.77 |

### 3.3. *Comparison of NHP&S mapped FAERS total and mAEOLUS*

Results from comparison of NHP&S mapped mAEOLUS with FAERS revealed that on average 39.11±11.37% more ADE records were retrieved using Custom MetaMap integrated with ingredient dictionary when compared to using OHDSI vocabulary alone. The numbers were comparatively lower for years 2013 (18.14%) and 2014 (13.27%). Figure 2 indicates the year-wise comparison of NHP&S associated ADE records retrieved from mAEOLUS and FAERS.

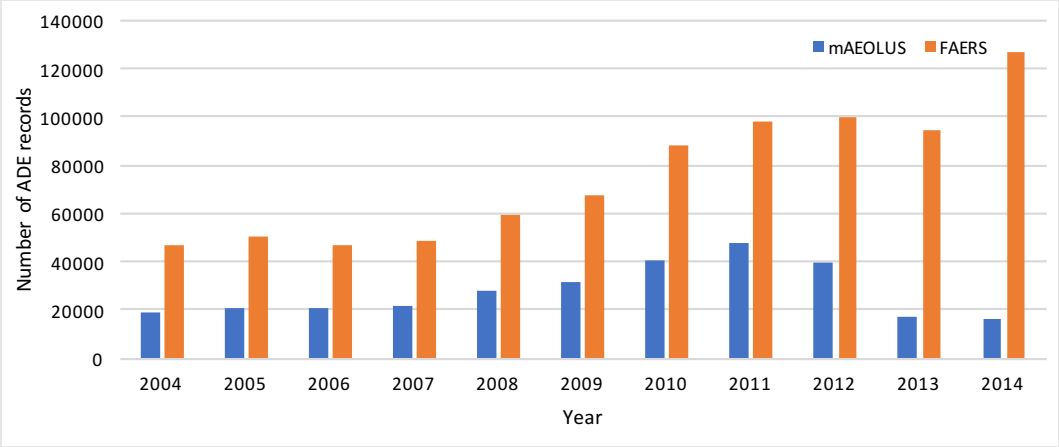


Figure 2. Comparison of ADE records identified from mAEOLUS and FAERS.

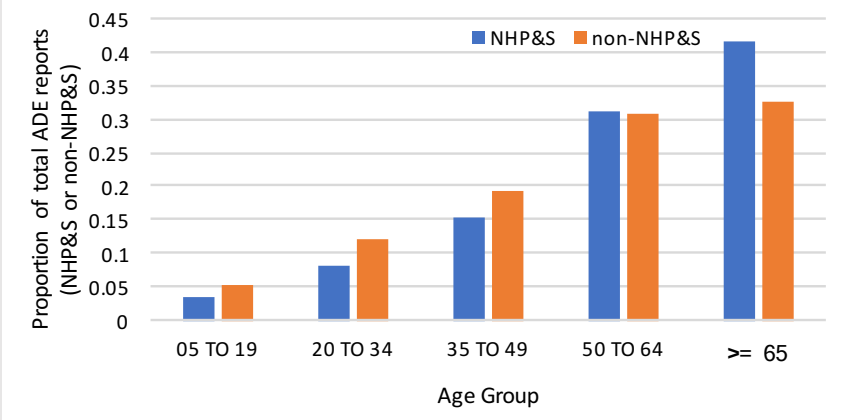


Figure 3. Comparison of NHP&S and non-NHP&S ADE report proportions stratified by age groups.

### 3.4. *Summary of NHP&S related reports in FAERS*

The NHP&S associated ADE records retrieved comprises of an average of 13.93±1.61% of total ADE records in FAERS every year from 2004 to 2014. The proportion of non-NHP&S ADE reports was higher in all age groups below 65. However, among the population with age group greater than or equal to 65, the proportion of NHP&S related reports was higher than non-NHP&S related reports (Figure 3). Figure 4 indicates the comparison of normalized values of NHP&S and non-NHP&S associated ADE report counts organized by SOCs. The top five SOC categories where the proportion of NHP&S is higher than non-NHP&S related ADE report counts were: (1) *Injury poisoning and procedural complications* (Inj&P); (2) *Gastrointestinal disorders* (Gastr); (3) *Infections and infestations* (Infec); (4) *Product issues* (Prod); and (5) *Metabolism and nutrition disorders* (Metab).

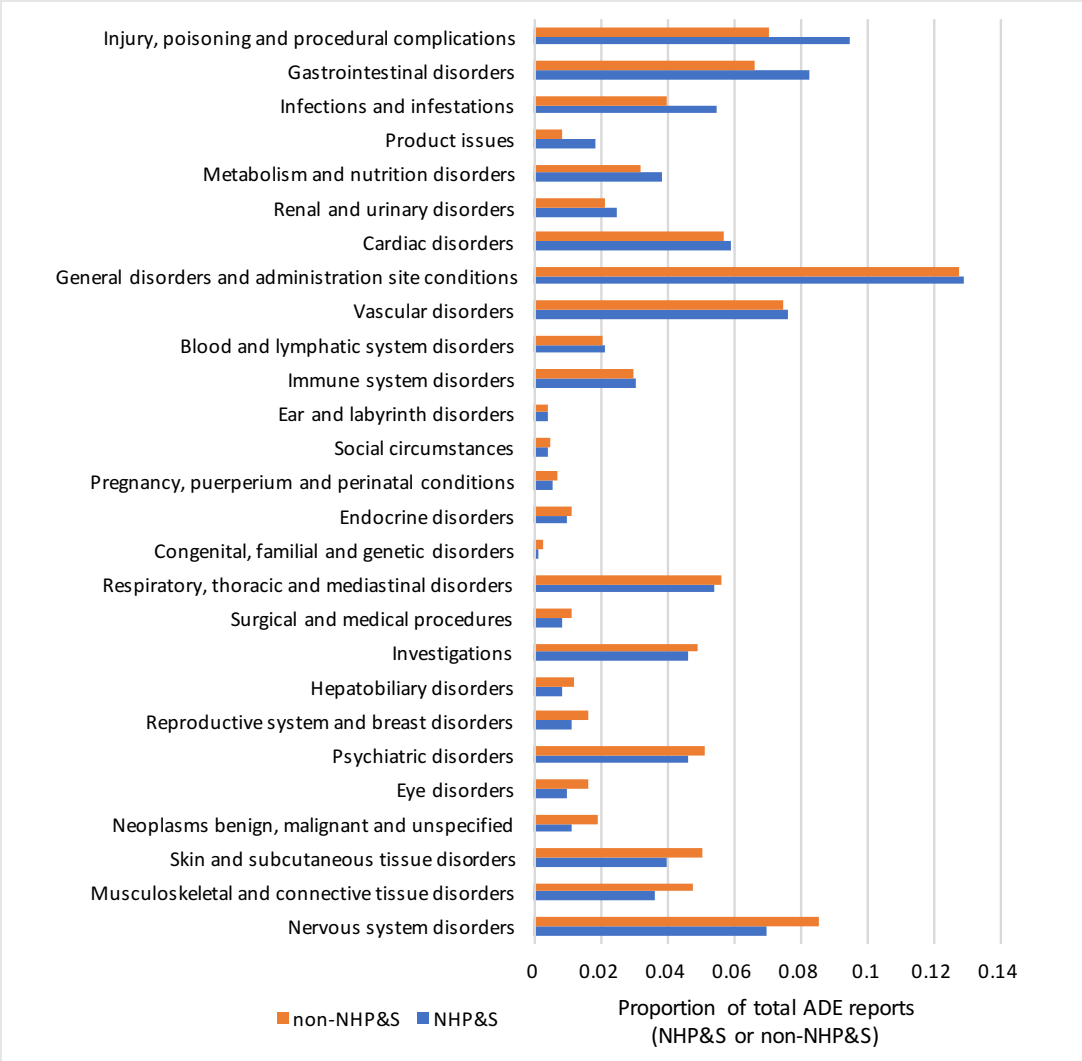


Figure 4. Comparison of proportions of NHP&S and non-NHP&S reports grouped by SOC.

## 4. Discussion

The biomedical domain is equipped with rich and multiple vocabulary sources and tools and techniques for concept recognition from unstructured data fields. Such resources play a key role in automation of cataloguing, indexing and retrieval of information of interest. Use of tools and techniques that provide scalable solution to analyze large amounts of data aid the discovery and generation of actionable outputs. Although there is significant amount of data publicly available in biomedical domain, the pipeline to attain insights from such data suffers from multiple hurdles. This in turn restricts the multi-disciplinary access and sharing of such data to those interested in translational research. The impediments in making the data accessible lies in data extraction, cleaning, standardization, and integration spanning multiple sources. Having these steps performed effectively may potentially facilitate design and execution of extensive data analysis plans. Community efforts such as the OHDSI focus on catering to such needs of researchers in biomedical community. However, the realm of NHP&S research lacks such resources which present a hurdle in pursuing data-driven investigations. This study explored the feasibility and utility of creating an NHP&S ingredient term thesaurus that could be leveraged by existing NLP tools for identifying relevant information embedded within biomedical knowledge sources.

The constructed NHP&S thesaurus for this study was a compilation of natural health products and dietary supplement ingredients from sources that have been either curated using evidence-based information (e.g., RxList, Medscape, Natural Medicines), reviewed by experts from FDA and the United States Pharmacopeia (e.g., SRS-UNII), issued a product license (e.g., LNHPD), or are/were available in the U.S. market (DSLD). A major challenge among the ingredient terminology sources is the lack of coverage of full set of synonyms, scientific names, and vernacular ("common") names[28]. Ambiguity of scientific names is another challenging aspect which requires close attention. Because many natural products are based on organism names, future work will include identifying natural product ingredient source organism names and gathering complete list of accepted scientific names, synonyms, and vernacular names. Similarly, for chemical dietary supplement ingredients, accepted IUPAC names, commonly used names, and abbreviations need to be included. In addition to the ingredient names, having commonly used product names in the thesaurus may improve the recall.

The results from evaluation suggests the need for development of NLP systems with enhanced mapping ability. The underreported nature adverse events related to DS, with only one in 100 being reported to FDA[29], accentuates the need for tools and approaches with higher sensitivity. Such tasks could benefit from the recent advancements being made in entity recognition from text sources, such as deep learning methods such as long-short term memory (LSTM)[30] or approaches combined with statistical word embeddings (LSTM-CRF)[31]. The comparison of results from mapping mAEOLUS with raw FAERS data with custom MetaMap shows recovery of additional ADE reports that were otherwise missed, potentially due to inadequate NHP&S ingredient term coverage within the current OHDSI vocabulary. The lower numbers for 2013 (18.14%) and 2014 (13.27%) could be due to less NHP&S records in missing AEOLUS rows or those already in the OHDSI vocabulary. Future work will be focused on grouping similar interventions associated with NHP&S. The incomplete grouping of entry terms within current thesaurus reflects higher number

of concepts. Efforts to expand this study will focus on manual curation to use relations to group terms into an ontological structure. Such step will result in fewer number of actual concepts/entries representing a compact NHP&S collection and will allow efficient categorization of their respective adverse events for calculation of signal disproportionality statistics. Such complete thesaurus of NHP&S would enable retrieval of relevant information from a variety of sources such as biomedical literature, clinical notes, online health forums, and social media. In addition to retrieval and dissemination of data in a standardized form, this effort will promote interoperability among traditionally disconnected data sources leading to generation of insights from more comprehensive analysis of data with limited risk of information loss.

NHP&S ADE reports may be important for analysis and detection of adverse event signals, both in terms of direct effects as well as interactions with pharmaceutical drugs. The proportion of adverse events related to NHP&S was higher in age group greater than or equal to 65 (senior citizens) compared non-NHP&S (Figure 3). This finding is consistent with the findings reported to the Special Committee on Aging, U.S. Senate[32]. Grouped by the top hierarchical structure of MedDRA, SOC, the proportion of injury and poisoning (*Inj&P*) related reports were higher in NHP&S group compared to non-NHP&S (Figure 4). The results presented here demonstrate the added potential for leveraging existing biomedical knowledge sources, such as FAERS, as a source for NHP&S knowledge.

This study highlights several challenges and opportunities in development of vocabulary resource and terminology mapping approaches for fostering advanced analytic investigations. A glimpse of utility of such resources in studying FAERS reports makes the case for investing the required time and effort in further enhancement to this infrastructure. Community wide efforts are required in this domain to make data accessible in standardized form in order to scale up to the methodological advances as exists in biomedical domain focused on drug-HOI associations.

## 5. Conclusion

This study developed a new NHP&S thesaurus for supporting the processing, identification, and standardization of relevant NHP&S data from existing digital resources. The application of the NHP&S thesaurus enabled a greater than 39% improvement in identifying NHP&S adverse events from the FAERS dataset. Such promising results suggest that there may be systematic approaches for identifying crucial NHP&S knowledge from existing biomedical data sources, and thus support overall curation efforts from complementary initiatives to develop community resources. Supplementary data are at: https://sites.google.com/a/brown.edu/phytokb/psb2018

## 6. Acknowledgement

This study was funded in part by NIH grants U54GM115677 and R01LM011963.

## References

1. Kocbek S, Groza T. Building a dictionary of lexical variants for human phenotype descriptors. In: Proceedings of the 15th Workshop on Biomedical Natural Language Processing. Berlin, Germany: Association for Computational Linguistics; 2016. p. 186–90.

2. Sarkar IN, Butte AJ, Lussier YA, Tarczy-Hornoch P, Ohno-Machado L. Translational bioinformatics: linking knowledge across biological and clinical realms. J Am Med Inform Assoc. 2011 Jul;18(4):354–7.
3. Altman RB. Translational bioinformatics: linking the molecular world to the clinical world. Clin Pharmacol Ther. 2012 Jun;91(6):994–1000.
4. Sarkar IN, Cantor MN, Gelman R, Hartel F, Lussier YA. Linking biomedical language information and knowledge resources: GO and UMLS. Pac Symp Biocomput. 2003;439–50.
5. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. Nucleic Acids Res. 2004 Jan 1;32(Database issue):D267–70.
6. Boyce RD, Ryan PB, Norén GN, Schuemie MJ, Reich C, Duke J, et al. Bridging islands of information to establish an integrated knowledge base of drugs and health outcomes of interest. Drug Saf. 2014 Aug;37(8):557–67.
7. 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.
8. Evaluation CFD, Research. Drug Approvals and Databases - FDA Adverse Event Reporting System (FAERS). [cited 2017 Jul 18]; Available from: https://www.fda.gov/drugs/informationondrugs/ucm135151.htm
9. Office USGA. Health Products for Seniors: “Anti-Aging” Products Pose Potential for Physical and Economic Harm. 2001 Sep 10 [cited 2017 Jul 7];(GAO-01-1129). Available from: http://www.gao.gov/products/gao-01-1129
10. 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 Oct 15;373(16):1531–40.
11. 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 Sep;116(9):1423–42.
12. 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.
13. 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 Nov 5;2015:1361–70.
14. Dietary Supplement Label Database (DSLD) [Internet]. [cited 2017 Jul 18]. Available from: https://dsld.nlm.nih.gov/dsld/index.jsp
15. Wang Y, Adam TJ, Zhang R. Term Coverage of Dietary Supplements Ingredients in Product Labels. AMIA Annu Symp Proc. 2016;2016:2053–61.
16. Aronson AR, Lang F-M. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc. 2010 May;17(3):229–36.
17. Licensed Natural Health Products Database (LNHPD) - Canada.ca [Internet]. Health Canada. 2007 [cited 2017 Jul 18]. 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

18. SRS-UNII. Substance Registration System - Unique Ingredient Identifier (UNII) [Internet]. Office of the Commissioner, Center for Biologics Evaluation and Research, Center for Veterinary Medicine, Center for Devices and Radiological Health, Center for Food Safety and Applied Nutrition, Center for Drug Evaluation and Research, National Center for Toxicological Research; [cited 2017 Jul 18]. Available from: https://www.fda.gov/forindustry/datastandards/substanceregistrationsystem-uniqueingredientidentifierunii/
19. RxList - The Internet Drug Index for prescription drugs, medications and pill identifier [Internet]. [cited 2017 Jul 18]. Available from: http://www.rxlist.com/script/main/hp.asp
20. Welcome to the Natural Medicines Research Collaboration [Internet]. [cited 2017 Jul 18]. Available from: https://naturalmedicines.therapeuticresearch.com/
21. Latest Medical News, Clinical Trials, Guidelines – Today on Medscape [Internet]. [cited 2017 Jul 18]. Available from: http://www.medscape.com/
22. Carter JS, Brown SH, Bauer BA, Elkin PL, Erlbaum MS, Froehling DA, et al. Categorical information in pharmaceutical terminologies. AMIA Annu Symp Proc. 2006;116–20.
23. Liu S, Ma W, Moore R, Ganesan V, Nelson S. RxNorm: prescription for electronic drug information exchange. IT Prof. 2005;7(5):17–23.
24. Medical Subject Headings - Home Page. 1999 Sep 1 [cited 2017 Jul 18]; Available from: https://www.nlm.nih.gov/mesh/
25. Evaluation CFD, Research. Drug Approvals and Databases - National Drug Code Directory. [cited 2017 Jul 19]; Available from: https://www.fda.gov/drugs/informationondrugs/ucm142438.htm
26. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proc AMIA Symp. 2001;17–21.
27. Brown EG, Wood L, Wood S. The medical dictionary for regulatory activities (MedDRA). Drug Saf. 1999 Feb;20(2):109–17.
28. Sharma V, Sarkar IN. Leveraging biodiversity knowledge for potential phyto-therapeutic applications. J Am Med Inform Assoc. 2013 Jul;20(4):668–79.
29. US Government Accountability Report [GAO]. Dietary Supplements: FDA Should Take Further Actions to Improve Oversight and Consumer Understanding. Vol. GAO-09-250. 2009.
30. Liu Z, Yang M, Wang X, Chen Q, Tang B, Wang Z, et al. Entity recognition from clinical texts via recurrent neural network. BMC Med Inform Decis Mak. 2017 Jul 5;17(Suppl 2):67.
31. Habibi M, Weber L, Neves M, Wiegandt DL, Leser U. Deep learning with word embeddings improves biomedical named entity recognition. Bioinformatics. 2017 Jul 15;33(14):i37–48.
32. Office USGA. Health Products for Seniors: Potential Harm From “Anti-Aging” Products. 2001 Sep 10 [cited 2017 Jul 20];(GAO-01-1139T). Available from: http://www.gao.gov/products/GAO-01-1139T