Identifying Supplement Use Within Clinical Notes: An Application of Natural Language Processing

Vivekanand Sharma, PhD, Indra Neil Sarkar, PhD, MLIS
Center for Biomedical Informatics, Brown University, Providence, Rhode Island

Abstract

Recent statistics indicate that the use of dietary supplements has increased over the years. Although being popular among consumers who use them for a variety of reasons, there have been limited clinical data-driven studies of the impact of dietary supplements on health outcomes. Challenges that impede such analyses in a comprehensive manner include either the sequestered nature of such data or their embedding within biomedical and clinical text. This study explored the feasibility to uncover patterns in the use of supplements, focusing on vitamin use among patients diagnosed with mental illness within patient records from the MIMIC-III database. The relevance of vitamin(s) was calculated at different levels of granularity and compared with association identified from Dietary Supplement Subset of MEDLINE. The results reveal insights into vitamin use for specific mental health related diagnosis and highlight challenges with identifying supplement information from clinical sources.

Introduction

The estimated sales of dietary supplements in the United States totaled $36.7 billion in 2014, out of which $14.3 billion (38.96%) included vitamin and mineral containing supplements. According to the 2016 annual survey on dietary supplements conducted on behalf of Council for Responsible Nutrition (CRN), about 97% of dietary supplement users take vitamins or minerals. Despite debate on their potential benefits, the choice to use vitamins may be guided by reasons including: overall health and wellness, enhance energy, fill in nutrient gaps, and immune health. The research domain focused on studying dietary supplements is fraught with challenges that include inadequate data resources, lack of suitable means to measure health outcomes, and rudimentary quantification of risks and benefits. In order to aid data-driven methodologies to gain insight about the use and impact of supplement use, acquisition of data about supplement usage from existing biomedical and clinical knowledge sources may be essential.

Patients records provide a rich source of information that can allow for analysis of real-world effectiveness and use patterns. They encompass reports of symptoms, findings from examinations, as well as diagnostic tests, diagnoses, prescribed drugs, and other interventions, along with additional information on facets such as family, social, and behavioral aspects. Such data are present in health records as either structured or unstructured fields. Patient records have been used to study patterns and trends in use of medications and their respective health outcomes. Although rich in information content, the cumulative growth of data and presence of free text elements poses significant challenges in attaining meaningful information. Automated acquisition of clinically relevant knowledge from health records may provide ways to address such issues allowing analyses and dispersal of insights in a timely and high-throughput manner. Approaches leveraging Natural Language Processing (NLP) pipelines for identification of required information from free text fields has shown promise (e.g., tobacco use information from social history and clinical notes).

Concurrently, the evolving clinically relevant knowledge with new discoveries in biomedical domain has the potential to alter clinical practice. However, such knowledge is buried in vast amounts of biomedical literature (e.g., articles indexed within MEDLINE). The insights gained in a timely manner from rich sources of information such as patient records or biomedical literature have the potential to provide the groundwork for decision support, quality assurance, and clinical informational needs. Text mining provides essential ways to address the issue of extracting relevant information from larger corpora in an automated manner. One of the primary tasks in development of such pipelines is recognition of biomedical entities embedded within the text. Leveraging the NLP based recognition of entities, several studies have attempted to identify related entities (e.g., drug-disease and gene-disease) using co-occurring concepts. Such studies have also shown potential in identification of herb-disease associations to cater specific informational needs. Systems such as SemRep and BioMedLEE have also been used for integrating co-occurrence based systems with semantic relation extraction system for literature-based discovery. The existing
NLP systems to recognize biomedical entities and relations have relied on rule-based or machine learning based approaches. More recent advancements in this area have leveraged deep learning for popular NLP tasks\textsuperscript{12}.

Significant advances have been made in acquisition and analyses of data related to the use of prescription drugs and associated health outcomes. However, such advancements have not yet been leveraged to its fullest potential in the domain of Natural Health Products and Supplements (NHP&S). The primary hurdle in conducting analyses of health outcomes associated with NHP&S results from data being either sequestered in domain specific sources lacking resources for interoperability or embedded within the general biomedical and clinical data sources\textsuperscript{17}. There are myriad challenges surrounding the tools and techniques to study NHP&S, including the lack of domain-specific vocabularies and inadequate term coverage by conventional biomedical resources. There is a need for development of tools, techniques, and resources for supporting the needs for making the data available in a standardized format that could facilitate the subsequent analyses\textsuperscript{14}.

Collating information on efficacy and safety of supplements may equip patient and healthcare providers with important knowledge to make informed choices, especially in light of debatable effects on health outcomes. This has been observed in case of existing understanding of the connection between dietary supplements and mental illness. With the introduction of psychiatric medications in the 1950s there has been a decline in interest in studies related to relevance of dietary nutrients for mental health\textsuperscript{15}. Among other supplements, the use of vitamins and minerals is more popular. While the use of this subset of supplements is popular, the pattern of use with regards to specific ailments requires further investigation. Studies have shown that low levels of vitamins can be related with mental illness, for example, association between low levels of folic acid with depressive symptomatology and poor response to antidepressant medications\textsuperscript{16}. Vitamin B6, B8, and B12 have been shown to reduce psychiatric symptoms and reduce the duration of illness\textsuperscript{4}. There is evidence in scientific literature that reflect the use of specific vitamins for treating specific psychiatric symptoms (e.g., use of Vitamin B1 to treat anxiety disorders including symptoms such as insomnia, nightmares, anorexia, nausea and vomiting\textsuperscript{3}). Similarly, it has been proposed that a subset of schizophrenia patients can respond to niacin augmentation therapy better than others\textsuperscript{17}. Such evidence suggest the utility in studying patterns of vitamin use and efficacy at different levels of diagnostic granularity.

This study examines the feasibility of mining relevant information about supplement use from patient records and provides a comparison with biomedical literature, focusing on a case study of vitamin use among patients diagnosed with mental illness. Associations related to vitamins were specifically calculated with respect to overall use of other supplements among patients with mental illness. A custom thesaurus integrated with an NLP tool was used to process the text for identification of NHP&S mentions. The feasibility and relevance of identified associations were examined by leveraging the granularity of International Classification of Diseases codes, Version 9, Clinical Modification (ICD-9-CM) and Clinical Classifications Software (CCS) Single-Level Diagnoses categories (CCS category). The results from this case study provide insight into potential improvements needed for identification and mapping approaches to support study of supplement use within clinical contexts. The results also reveal some early perspectives on vitamin use for specific diagnoses within the larger category of mental disorders and suggest that this area of research warrants a more in-depth clinical data-driven evaluation of health outcomes.

**Materials and Methods**

The primary goal of this study was to provide a comparative view outlining the use of vitamins with respect to other NHP&S among patients with diagnoses falling within the ICD-9-CM category\textsuperscript{18} of “Mental Disorders (290-319)”. A custom thesaurus of NHP&S terms integrated with biomedical NLP tool MetaMap\textsuperscript{19} (built using the MetaMap Data File Builder suite\textsuperscript{20}) was used to facilitate extraction of relevant information from the data sources of interest. Standard MetaMap was used to identify diagnoses terms from MEDLINE text and further mapped to ICD-9-CM\textsuperscript{21}. Associations between NHP&S terms and mental health related diagnoses terms were calculated at two levels: (1) ICD-9-CM codes; (2) CCS Single-Level Diagnoses categories (“CCS category”)\textsuperscript{22}. A general overview of the approach is depicted in Figure 1.

**Data sources and filters used for the study**

Data from two sources were considered for this study: De-identified patient records from the Medical Information Mart for Intensive Care III (MIMIC-III)\textsuperscript{23} database and biomedical literature indexed in MEDLINE\textsuperscript{24}.

**MIMIC-III**: Provided by the Beth Israel Deaconess Medical Center, MIMIC-III is a publicly available repository containing de-identified data from over 40,000 critical care patients ranging from year 2001 to 2012. Such a resource can be leveraged to support studies including epidemiology, clinical decision-rule improvement, and
electronic tool development. The data contained within MIMIC-III spans demographics, vital sign measurements, laboratory tests, procedures, medications, and notes and reports provided by the healthcare provider. These data elements are organized into several tables in relational form. For the purpose of this study, two of the tables were of interest: (1) DIAGNOSES_ICD: Provides a list of diagnoses for a given patient identifier coded using the ICD-9-CM; (2) NOTEEVENTS: Contains notes from providers in several categories such as discharge summary, nursing, nutrition, physician, and radiology. Focusing on diagnoses related to mental health disorders, a SQL query was used to identify data (diagnoses and text notes) associated with patients with at least one diagnosis from the ICD-9-CM category, “Mental Disorders (290-319)”.

**Figure 1: Study Overview.** Patient records from associated with ICD-9-CM category “Mental Disorders” were processed to identify mentions of dietary supplements. Corresponding articles related to the identified ICD-9-CM codes were identified from MEDLINE (Dietary Supplement Subset). Associations between supplements and diagnoses (ICD-9-CM codes or CCS Categories) were calculated and results were presented using vitamin use as a case study.

**MEDLINE:** Provided by the National Library of Medicine (NLM)\(^25\), MEDLINE is one of the primary sources of biomedical literature and references. It contains more than 24 million references from journals selected based on a comprehensive selection process. The article entries within MEDLINE are indexed using a controlled vocabulary, the Medical Subject Headings (MeSH)\(^26\). Focused specifically on articles related to the use of NHP&S by patients diagnosed with mental health issues, the search relied on a subset of literature related to dietary supplements. This “Dietary Supplement Subset”\(^27\) is provided by the partnership of the Office of Dietary Supplements\(^28\) and the NLM to restrict the search results to articles related to a broad spectrum of literature including vitamins, minerals, phytochemicals, ergogenic, botanical, and herbal supplements in human nutrition and animal models. Using this subset as a primary filter, an additional filter for mental health disorders was applied using MeSH descriptor “Mental Disorders (F03)” (Query: "Mental Disorders[mh] AND dietsuppl[sb]"). An additional article set where mental disorders was the central theme was also identified (Query: "Mental Disorders[majr] AND dietsuppl[sb]") and retained for comparison.

**Custom thesaurus of NHP&S terms.** A custom thesaurus was used for this study for domain specific information acquisition (created as a part of previous study\(^14\)). This thesaurus was built to incorporate NHP&S terms from sources that offer reliable and comprehensive coverage of terms, synonyms, and variants. The selection of terms was based on sources that aim to provide evidence-based information to healthcare providers and general public. The primary selection of terms was from six sources: LNHPD\(^29\), DSLD\(^30\), SRS-UNII\(^31\), RxList\(^32\), Natural Medicines\(^33\), and Medscape\(^34\) were further mapped to UMLS generally and specifically to NDF-RT\(^35\), RxNorm\(^36\), MeSH\(^37\). Similar strings, synonyms, and variants across different sources were grouped together into concepts and assigned unique identifiers. 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 the NLP tool MetaMap (“Custom MetaMap”).
Processing of MIMIC notes. The notes identified from MIMIC-III database table NOTEEVENTS were stored as separate text files for each patient. These text files were processed using Custom MetaMap and the machine output was parsed using a Julia program retaining mappings with a perfect score of 1000.

Processing of MEDLINE abstracts. The titles and abstracts were extracted from the MEDLINE subset in XML format. The resulting set of text was processed separately using MetaMap Java API with: (1) Custom MetaMap: To identify NHP&S mappings (with score of 1000); (2) Default MetaMap: To identify mappings to diagnoses terms. The mappings from text to ICD-9-CM codes were performed using a two-fold approach: (a) Direct Mapping: The default MetaMap processing was restricted to ICD-9-CM (-q – ICD9CM). The UMLS CUIs identified with perfect mapping scores were used to directly query UMLS MRCONSO table to retain a corresponding ICD-9-CM code; and (b) Inferred Mapping: The text was processed using default MetaMap and the output was parsed to retain UMLS CUIs (with mapping score of 1000) of semantic types belonging to the UMLS semantic group “Disorders”. The UMLS CUIs were queried against UMLS MRCONSO to identify direct mappings. If no direct mappings were found, a recursive search query (five iterations) was performed on UMLS MRREL table (including relationship type ‘isa’) to identify mappings to ICD-9-CM. This search approach followed UMLS CUIs within MRREL table to identify entities that are related by relationship type ‘isa’ until a CUI was identified that could be directly mapped to ICD-9-CM. The identification of ‘isa’ related CUIs and checking for mapping to ICD-9-CM codes was pursued up to five iterations. Following the processing of text and identifying NHP&S terms and corresponding ICD-9-CM codes, those records that had ICD-9-CM codes identified from list of diagnoses from MIMIC dataset within the ICD category “Mental Disorders (290-319)” were retained. For all ICD-9-CM codes, their corresponding CCS Single-Level Diagnoses categories were identified and used for further analyses. The CCS groups more than 14,000 ICD-9-CM diagnosis codes into a smaller number of clinically meaningful categories. The categories include classifications for Mental Health and Substance Abuse (CCS-MHSA)\(^2\).

Associations among NHP&S terms and ICD9CM diagnoses. For both datasets, associations between NHP&S terms and diagnoses terms (both individual ICD-9-CM codes and CCS categories) were calculated using Odds Ratio (OR) with a threshold of at least five or more subjects with a given ICD-9-CM diagnosis using a particular vitamin. Associations with OR greater than one, having 95% Confidence Interval (95% CI) that did not include one were retained for further comparison between the MIMIC and MEDLINE datasets.

Results
Top ICD-9-CM codes with respect to vitamin use. A total of 13,400 subjects were identified from MIMIC-III data having at least one diagnoses falling in the “Mental Disorder (290-319)” category. Out of 13,400 subjects, 3248 (24.24%) use one or more vitamins. The demographics of study population and vitamin users is provided in Table 1.

### Table 1: Demographics of subject population

|                | Vitamin Users | All Subjects | Percent |
|----------------|---------------|--------------|---------|
| **Sex**        |               |              |         |
| Male           | 1799          | 7414         | 24.26   |
| Female         | 1455          | 5986         | 24.31   |
| **Age Groups** |               |              |         |
| > 5            | 1             | 7            | 14.29   |
| 05-19          | 6             | 74           | 8.11    |
| 20-34          | 177           | 1149         | 15.4    |
| 35-49          | 616           | 2309         | 26.68   |
| 50-64          | 1003          | 4009         | 25.02   |
| ≥65            | 1451          | 5852         | 24.79   |
| **Ethnicity**  |               |              |         |
| White          | 2407          | 9742         | 24.71   |
| Black          | 293           | 1075         | 27.26   |
| Unknown        | 190           | 966          | 19.67   |
| Hispanic       | 93            | 397          | 23.43   |
| Unknown        | 68            | 304          | 22.37   |

### Table 2: Vitamin use by subjects categorized by ICD-9-CM (threshold 10%)

| ICD-9-CM | Description                | Subjects | Percent |
|----------|----------------------------|----------|---------|
| 311      | Depressive disorder NEC    | 815      | 25.09   |
| 305.1    | Tobacco use disorder       | 640      | 19.70   |
| 291.81   | Alcohol withdrawal         | 404      | 12.44   |
| 293.0    | Acute delirium             | 387      | 11.92   |
| 303.91   | Alcoh dep NEC/NOS-contin   | 358      | 11.02   |
| 300.00   | Anxiety state NOS          | 336      | 10.35   |
The top ICD-9-CM diagnoses codes for which vitamin use was prevalent (with a threshold of 10% of total vitamin users) were: (1) Depressive disorder NEC; (2) Tobacco use disorder; (3) Alcohol withdrawal; (4) Acute delirium; (5) Alcohol dependence; and (6) Anxiety state NOS. The count and their respective percentages are listed in Table 2.

**Distribution across CCS Single-Level Diagnoses Categories.** The ICD-9-CM codes identified from subjects using one or more vitamins were mapped to 14 CCS categories. The top CCS categories (threshold of 10%) with higher number of vitamin users were (CCS category code in parenthesis): (1) Alcohol-related disorders (660); (2) Mood disorders (657); (3) Delirium, dementia, and amnestic and other cognitive disorders (653); (4) Screening and history of mental health and substance abuse codes (663); (5) Substance-related disorders (661); and (6) Anxiety disorders (651). The distribution of the total of 3244 subjects using vitamins with at least one diagnoses in mental disorder category is listed in Table 3.

### Table 3: Count and percentage of subjects using vitamins categorized by CCS Single-Level diagnosis categories

| CCS Code | CCS Category Name                                      | Subjects | Percent |
|----------|-------------------------------------------------------|----------|---------|
| 650      | Adjustment disorders                                   | 35       | 01.08   |
| 651      | Anxiety disorders                                      | 433      | 13.33   |
| 652      | Attention-deficit/conduct/disruptive behavior disorders | 25       | 00.77   |
| 653      | Delirium, dementia, and amnestic and other cognitive disorders | 853   | 26.26   |
| 654      | Developmental disorders                                | 96       | 02.96   |
| 655      | Disorders usually diagnosed in infancy, childhood, or adolescence | 5    | 00.15   |
| 656      | Impulse control disorders, NEC                         | 2        | 00.06   |
| 657      | Mood disorders                                         | 1238     | 38.12   |
| 658      | Personality disorders                                  | 38       | 01.17   |
| 659      | Schizophrenia and other psychotic disorders            | 148      | 04.56   |
| 660      | Alcohol-related disorders                              | 1271     | 39.13   |
| 661      | Substance-related disorders                            | 469      | 14.44   |
| 663      | Screening and history of mental health and substance abuse codes | 640   | 19.70   |
| 670      | Miscellaneous mental health disorders                  | 65       | 02.00   |

**Table 4: Proportion of subjects using vitamins across ICD9CM diagnoses codes (threshold 10%) within top CCS categories**

| CCS Code | ICD-9-CM | ICD-9-CM description | Count | Percent |
|----------|----------|-----------------------|-------|---------|
| 660      | 291.81   | Alcohol withdrawal    | 404   | 12.44   |
|          | 303.91   | Alcohol dependence    | 358   | 11.02   |
| 657      | 311      | Depressive disorder NEC| 815   | 25.09   |
| 653      | 293.0    | Acute delirium        | 387   | 11.92   |
| 663      | 305.1    | Tobacco use disorder  | 640   | 19.70   |
| 661      | NA       | NA                    | NA    | NA      |
| 651      | 300.00   | Anxiety state NOS     | 336   | 10.35   |

**ICD-9-CM codes within CCS categories.** Keeping a threshold proportion of 10%, the top ICD-9-CM diagnoses among vitamin users for CCS category “Alcohol-related disorders (660)” was Alcohol withdrawal (291.81) and Alcohol dependence (303.91). Similarly, for “Mood disorders (657)”, the top ICD-9-CM code was Depressive disorder NEC (311). For CCS category “Delirium, dementia, and amnestic and other cognitive disorders (653)” the top ICD-9-CM code was Acute delirium (293.0). For “Screening and history of mental health and substance abuse codes (663)” the top ICD9CM code was Tobacco use disorder (3051) and for “Anxiety disorders (651)” the ICD9CM code was Anxiety state NOS (300.00). A summary of the proportions of top ICD-9-CM diagnoses within top CCS categories is listed in Table 4. A list of vitamins significantly associated (as identified from MIMIC-III and MEDLINE dataset) with the top ICD-9-CM diagnoses codes mentioned above is presented in Table 5.

**Vitamin use associations among CCS categories.** Significant associations of vitamin use and ICD-9-CM diagnoses were identified for 11 out of the initially identified 14 CCS categories. These associations were either from MIMIC-III dataset or MEDLINE dataset (using direct UMLS CUI to ICD-9-CM mapping or inferred UMLS CUI to ICD-9-CM mapping). For the MEDLINE dataset, a higher number of associations were identified when the inferred
mapping (via UMLS MRREL relationship type ‘isa’) was used. Common associations among MIMIC-III and MEDLINE dataset were identified for only three CCS categories: (1) Delirium, dementia, and amnestic and other cognitive disorders (653): Vitamin E and B12; (2) Alcohol-related disorders (660): Vitamin B1; and (3) Miscellaneous mental health disorders (670): Vitamin D. The results from identification of associations using odds ratio across different CCS categories for MIMIC and MEDLINE dataset is summarized in Table 6.

Table 5: Vitamin associated with ICD-9-CM codes from Table 4

| ICD-9-CM | ICD-9-CM description | MIMIC-III | MEDLINE |
|----------|----------------------|-----------|---------|
|          | Vitamin              | OR        | 95% CI  | p-value | OR         | 95% CI  | p-value |
| 291.81   | Alcohol withdrawal   | Multivitamin 9.71 3.31-28.49 p < 0.0001 | Vitamin B1 6.42 5.50-7.49 p < 0.0001 | 6.19 2.39-16.02 p = 0.0002 |
|          |                      | Vitamin B9 5.22 4.50-6.05 p < 0.0001 | Vitamins 3.12 1.73-5.63 p = 0.0002 |             |
|          |                      | Vitamin B12 2.69 1.33-5.43 p = 0.0057 |             |             |          |
| 303.91   | Alcohol dependence   | Vitamin B1 7.10 4.47-11.29 p < 0.0001 | Vitamin B9 4.41 3.77-5.16 p < 0.0001 |             |
|          |                      | Vitamins 3.73 2.14-6.51 p < 0.0001 | Multivitamins 3.69 2.96-4.60 p < 0.0001 |             |
|          |                      | Vitamin B12 2.41 1.15-5.04 p = 0.0194 | Vitamin B12 2.01 1.35-2.99 p = 0.0006 |             |
| 311      | Depressive disorder  | Vitamin D2 2.08 1.56-2.78 p < 0.0001 | Vitamin D3 1.54 1.32-1.80 p < 0.0001 | Vitamin B9 1.60 1.15-2.22 p = 0.0050 |
| NEC      | NEC                  | Vitamin B12 1.41 1.07-1.85 p = 0.0143 |             | Vitamin D 1.51 1.02-2.23 p = 0.0370 |          |
| 293.0    | Acute delirium       | Vitamin K 1.92 1.43-2.58 p < 0.0001 | Vitamin B12 1.87 1.03-3.41 p = 0.0411 |             |
|          |                      | Vitamin B12 1.56 1.03-2.36 p = 0.0339 | Vitamin D3 1.41 1.14-1.74 p = 0.0014 |             |
| 305.1    | Tobacco use disorder | Vitamin D 1.96 1.27-3.03 p = 0.0026 | Vitamin D2 1.52 1.01-2.30 p = 0.0469 |             |
|          |                      | Vitamin D3 1.33 1.07-1.65 p = 0.0088 | Vitamin B12 1.48 1.04-2.11 p = 0.0286 |             |
| 300.00   | Anxiety state NOS    | Vitamin B12 1.48 1.04-2.11 p = 0.0286 |             |             |          |

Discussion

There is a continuous increase in spending on and use of dietary supplements including vitamins which may be unsupervised or prescribed by healthcare provider. In light of such growth in use of supplements, it becomes imperative to systematically study and analyze the associated health outcomes including both aspects whether there is actually any benefit and if there are any risks involved. Towards supporting such ventures to examine the efficacy and safety, extracting information from existing biomedical literature and health records may provide valuable insights. Although the importance of such sources is acknowledged, the magnitude of available data is increasing. Automated methods are therefore increasingly required for efficient cataloguing, indexing and retrieval of relevant information. Such pipelines will be essential to cater the needs of clinicians and patients to make informed choices. Though a case study of vitamin supplement use compared to other supplements among patients with mental illness, patterns in use and associations with more specific (granular) diagnoses was examined here from biomedical literature indexed in MEDLINE and patient records in MIMIC-III database. Using ICD-9-CM codes and categorization in conjunction to the CCS categories associated with mental disorders, both a general and granular account is illustrated.

Within MIMIC-III, this study indicates that a sizeable portion (24.24%) of patients with diagnoses falling in mental disorder category use a range of vitamins. The use by specific diagnosis (on the basis of ICD-9-CM codes) show higher use by patients with depressive disorder, tobacco use disorder, alcohol-related disorders, acute delirium and anxiety compared to other diagnoses (Table 2). A relatively broader picture is provided in Table 3 from where CCS categories with comparatively higher proportions of use is reflected. The highest proportion of use is by patients with diagnosis in the category of “Alcohol-related disorders” followed by “Mood disorders”, “Delirium, dementia,
Table 6: Associations across different CCS categories for MIMIC and MEDLINE dataset as reflected from Odds Ratio (OR) and 95% Confidence Interval (CI)

| CCS | MIMIC | MEDLINE-DIRECT ICD-9-CM MAPPING | MEDLINE-INFERRED ICD-9-CM MAPPING | COMMON |
|-----|-------|-------------------------------|---------------------------------|--------|
|     | Vitamin | OR    | 95% CI       | Vitamin | OR    | 95% CI       | Vitamin | OR    | 95% CI       |
| 650 | Vitamin D2 | 3.96   | 1.82-8.59     | NA      | NA    | NA          | Vitamin B1<sup>1,2</sup> | 4.16   | 2.21-7.83     | NA    |
| 651 | Vitamin D2 | 1.58   | 1.04-2.41     | NA      | NA    | NA          | Vitamin E<sup>1,2</sup> | 4.24   | 2.73-6.58     | Vitamin E<sup>1,2</sup> |
|     | Vitamin B12 | 1.45   | 1.05-2.00     | NA      | NA    | NA          | Vitamin B12<sup>1,2</sup> | 2.85   | 1.93-4.22     | NA    |
|     | Vitamin D3 | 1.44   | 1.09-1.91     | NA      | NA    | NA          | Vitamin B12<sup>1,2</sup> | 2.01   | 1.19-3.40     | NA    |
|     | Vitamin D3 | 1.25   | 1.02-1.53     | NA      | NA    | NA          | Vitamin B12<sup>1,2</sup> | 2.81   | 1.64-3.14     | Vitamin B12 |
| 653 | Vitamin E  | 3.22   | 1.73-6.00     | Vitamin E<sup>1,2</sup> | 4.24   | 2.73-6.58     | Vitamin B12<sup>1,2</sup> | 2.81   | 1.64-3.14     | NA    |
|     | Vitamin D  | 1.45   | 1.15-1.83     | NA      | NA    | NA          | Vitamin B12<sup>1,2</sup> | 2.81   | 1.64-3.14     | NA    |
|     | Vitamin D3 | 1.67   | 1.44-1.94     | NA      | NA    | NA          | Vitamin B12<sup>1,2</sup> | 2.81   | 1.64-3.14     | NA    |
|     | Vitamin B12 | 1.47   | 1.13-1.92     | NA      | NA    | NA          | Vitamin B12<sup>1,2</sup> | 2.81   | 1.64-3.14     | NA    |
|     | Vitamin K  | 1.57   | 1.24-2.00     | NA      | NA    | NA          | Vitamin B12<sup>1,2</sup> | 2.81   | 1.64-3.14     | NA    |
| 654 | Vitamin D3 | 1.89   | 1.35-2.65     | Vitamin B12<sup>1</sup> | 10.68 | 3.73-30.59   | Vitamin B12<sup>1,2</sup> | 4.42   | 1.91-10.25    | NA    |
| 655 | NA        | NA     | NA            | NA      | NA    | NA          | Vitamin B12<sup>1,2</sup> | 2.81   | 1.64-3.14     | NA    |
| 657 | Vitamin D2 | 1.97   | 1.45-2.67     | NA      | NA    | NA          | Vitamin B9<sup>1,2</sup> | 1.38   | 1.02-1.87     | NA    |
|     | Vitamin D3 | 1.41   | 1.23-1.62     | NA      | NA    | NA          | Vitamin B9<sup>1,2</sup> | 1.38   | 1.02-1.87     | NA    |
| 658 | Multivitamins | 2.02  | 1.08-3.77     | NA      | NA    | NA          | Vitamin B9<sup>1,2</sup> | 1.38   | 1.02-1.87     | NA    |
| 659 | Multivitamins | 1.42  | 1.02-1.98     | Vitamin B3<sup>1</sup> | 3.21   | 1.16-8.88   | Vitamin B9<sup>1,2</sup> | 2.49   | 1.28-4.86     | NA    |
|     | Multivitamins | 3.06  | 2.41-3.88     | Vitamin B3<sup>1</sup> | 3.21   | 1.16-8.88   | Vitamin B9<sup>1,2</sup> | 2.49   | 1.28-4.86     | NA    |
| 660 | Vitamin B1  | 4.38   | 2.97-6.47     | Vitamin B1<sup>1,2</sup> | 18.02 | 7.32-44.37  | Vitamin B1<sup>1,2</sup> | 7.66   | 4.10-14.32    | Vitamin B1 |
|     | Vitamin K  | 1.68   | 1.26-2.24     | NA      | NA    | NA          | Vitamin B1<sup>1,2</sup> | 7.66   | 4.10-14.32    | NA    |
|     | Vitamin B9  | 3.26   | 3.03-3.73     | NA      | NA    | NA          | Vitamin B1<sup>1,2</sup> | 7.66   | 4.10-14.32    | NA    |
| 661 | Vitamin B1  | 1.54   | 1.31-1.81     | NA      | NA    | NA          | Vitamin B1<sup>1,2</sup> | 7.66   | 4.10-14.32    | NA    |
|     | Vitamin B9  | 1.48   | 1.28-1.72     | NA      | NA    | NA          | Vitamin B1<sup>1,2</sup> | 7.66   | 4.10-14.32    | NA    |
| 670 | Multivitamins | 2.17  | 1.10-4.28     | Vitamin D<sup>1,2</sup> | 2.89   | 1.36-6.14   | Vitamin D<sup>1,2</sup> | 2.89   | 1.36-6.14     | NA    |
|     | Vitamin D  | 1.94   | 1.08-3.50     | NA      | NA    | NA          | Vitamin D<sup>1,2</sup> | 2.89   | 1.36-6.14     | NA    |
|     | Vitamin B1  | 3.47   | 1.39-8.67     | NA      | NA    | NA          | Vitamin D<sup>1,2</sup> | 2.89   | 1.36-6.14     | NA    |

<sup>1</sup>MEDLINE query "Mental Disorders[mh] AND dietsuppl[sb];<sup>2</sup>MEDLINE query "Mental Disorders[majr] AND dietsuppl[sb]; The OR scores provided in table indicate the results obtained from "Mental Disorders[mh] AND dietsuppl[sb]; p-values calculated for all associations was < 0.05.}

with the proportions of vitamin use is provided in Table 4. Associations of several vitamins were identified from MIMIC-III dataset (Table 5) for the identified ICD-9-CM codes. Among other vitamins, Vitamin B1 shows significant association for alcohol withdrawal diagnosis. This observation was corroborated by analyses from MEDLINE data. Vitamin B1 (Thiamine) has been studied extensively for treatment of patients with alcohol dependence<sup>18</sup>. The MEDLINE query for this association ("Alcohol-Related Disorders[mh] AND "thiamine"[mh]) results in 838 articles. A limitation that needs to be considered in the inference pipeline is that the mapping of MEDLINE text with ICD-9-CM codes is not a perfect art. As reflected from Table 6, additional associations can be identified from MEDLINE when using an inferred mapping approach. Unlike the clinical notes and diagnoses mentions, the mentions of vitamin use for a given disorder may not be very specific, resulting in missing of some associations. Additionally, the possibility of information loss as a result of incomplete collection of articles indexed as “Dietary Supplement Subset” cannot be ignored. Significant association of Vitamin B9 for depressive disorder was identified from MEDLINE which was missing from MIMIC-III dataset. It has been shown that Vitamin B9 exerts antidepressant effects by upregulating brain-derived neurotrophic factor and glutamate receptor 1 expression...
in brain\textsuperscript{39} The use of Vitamin E and B12 for the CCS category “Delirium, dementia, and amnestic and other cognitive disorders (653)” as reflected from MIMIC-III data is corroborated by MEDLINE. There is considerable amount of literature that reflects Vitamin B3 association with schizophrenia treatment. The analysis in this study from MEDLINE dataset also reflects significant association between Vitamin B3 and CCS category “Schizophrenia and other psychotic disorders (659)”. However, such association was not recovered from MIMIC dataset. With respect to specific Vitamin D associations, a potential confounding factor might be the diagnostic test referred to by the physician which may have been picked up by the NLP system. The patterns of use of supplements in general may not completely reflect and align with evidence-based associations identified due to mostly lack of physician supervised nature of intake\textsuperscript{40}.

This study highlights several challenges and opportunities in mining patient data and biomedical literature for understanding the patterns of use of supplements. One specific aspect is having a supplement thesaurus attached to an NLP system that can provide full coverage and efficient mapping of supplement terms. Inadequate coverage of supplement terms has been shown to be a major issue in this area\textsuperscript{39}. Issues related to processing of clinical notes related to incomplete coverage of supplements has also been highlighted\textsuperscript{42}. Zhang \textit{et al.} point out the importance of clinical notes in mining and assessment of clinical effects of supplement interventions and also convey the difficulties in documentation as a results of gap between supplement term list and standard biomedical terminologies. Additional challenges in such analyses lies in annotation of MEDLINE text with more specific diagnoses codes that may enhance the retrieval of key associations.

Although the custom NLP pipeline for identification of NHP&S terms used in this feasibility study was evaluated in a prior study with FDA Adverse Event Reporting System (FAERS) dataset\textsuperscript{13} (Precision: 0.93; Recall: 0.66; and F-score: 0.77), there is a need to further evaluate the utility of the system within a clinical context. Nonetheless, preliminary insights can be derived from existing studies that provided evaluation of efficacy of MetaMap on clinical notes, as demonstrated here. The focus of this study was on whether supplement information could be identified from within clinical narratives; however, future work is needed to assess the clinical validity or utility of the identified correlations. Additionally, further improvements in identification and extraction of entities of interest could be attained using recent advances in NLP (e.g., leveraging deep learning approaches). Mapping of text from MEDLINE to ICD-9-CM codes is another challenging aspect that needs further improvement and evaluation. In order to retain the granularity aspect of associations, the approach followed in this study attempted to infer mappings by identifying related UMLS concepts rather than traversing the hierarchy for consolidation of ICD-9-CM codes from same higher level category. Manual inspection of a portion of EHR data from MIMIC-III reflects that most mentions of vitamins and other supplements are included under the section heading “Medications on Admission” and some or all of them are continued at the time of discharge as indicated in the section “Discharge Medications”. Extracting and restricting this study to specific sections on medications and diagnoses at the time of admission may potentially mask the confounding factor resulting from mentions in other sections of the clinical note not indicative of supplement use by the patient. Such an approach may portray more conclusive results by reducing biases associated with ICU-only nature of MIMIC-III dataset. The interpretation of results from this preliminary does require some scrutiny; however, this was an exploratory proof of concept study that suggests further data-driven investigation into patterns of vitamin use. Future work will be aimed at exploring such patterns of use in more general EHR data from more than one sources.

Conclusion

Published studies have suggested an association between vitamins and mental health disorders; however, the benefits associated with more granular diagnoses still needs to be analyzed. Development of automated methods for mining and evaluation of such information from clinical notes and biomedical literature may be valuable in terms of providing noticeable leads. Through a case study, focused on vitamin use in the context of mental health conditions, we attempted to examine the feasibility of identifying such associations with respect to other dietary supplements at various levels of granularity. Several promising leads were highlighted that suggest potential vitamin use for mental health related diagnosis. Additionally, this feasibility study reveals the potential for adapting existing biomedical tools and resources for effective mapping and identification of supplement information from clinical text.

Acknowledgements

We thank Dr. Paul Stey, Dr. Isabel Restrepo, Ashley Lee, and Dr. Elizabeth Chen for useful discussions and providing essential technical guidance with analyzing the MIMIC-III database. This work was funded in part by
References

1. Office of Dietary Supplements - Multivitamin/mineral Supplements [Internet]. [cited 2017 Sep 14]. Available from: https://ods.od.nih.gov/factsheets/MVMS-HealthProfessional/

2. CRN 2016 Annual Survey on Dietary Supplements | Council for Responsible Nutrition [Internet]. [cited 2017 Sep 14]. Available from: https://www.crnusa.org/resources/crn-2016-annual-survey-dietary-supplements

3. Guallar E, Stranges S, Mulrow C, Appel LJ, Miller ER 3rd. Enough is enough: Stop wasting money on vitamin and mineral supplements. Ann Intern Med. 2013 Dec 17;159(12):850–1.

4. Leahy LG. Vitamin B Supplementation: What’s the Right Choice for Your Patients? J Psychosoc Nurs Ment Health Serv. 2017 Jul 1;55(7):7–11.

5. Schneeweiss S, Avorn J. A review of uses of health care utilization databases for epidemiologic research on therapeutics. J Clin Epidemiol. 2005 Apr;58(4):323–37.

6. Chen ES, Hripcsak G, Xu H, Markatou M, Friedman C. Automated acquisition of disease drug knowledge from biomedical and clinical documents: an initial study. J Am Med Inform Assoc. 2008 Jan;15(1):87–98.

7. Wang Y, Chen ES, Pakhomov S, Lindemann E, Melton GB. Investigating Longitudinal Tobacco Use Information from Social History and Clinical Notes in the Electronic Health Record. AMIA Annu Symp Proc. 2016;2016:1209–18.

8. Frijters R, van Vugt M, Smeets R, van Schaik R, van Vlieg J, Alkema W. Literature mining for the discovery of hidden connections between drugs, genes and diseases. PLoS Comput Biol [Internet]. 2010 Sep 23;6(9). Available from: http://dx.doi.org/10.1371/journal.pcbi.1000943

9. Sharma V, Law W, Balick MJ, Sarkar IN. Identifying Plant-Human Disease Associations in Biomedical Literature: A Case Study. AMIA Jt Summits Transl Sci Proc. 2016 Jul 19;2016:84–93.

10. Sharma V, Sarkar IN. Leveraging concept-based approaches to identify potential phyto-therapies. J Biomed Inform. 2013 Aug;46(4):602–14.

11. Hristovski D, Friedman C, Rindflesch TC, Peterlin B. Exploiting semantic relations for literature-based discovery. AMIA Annu Symp Proc. 2006;349–53.

12. 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.

13. Sharma V, Sarkar IN. Leveraging biodiversity knowledge for potential phyto-therapeutic applications. J Am Med Inform Assoc. 2013 Jul;20(4):668–79.

14. Sharma V, Sarkar IN. Identifying natural health product and dietary supplement information within adverse event reporting systems. Pac Symp Biocomput. 2018;23:268–79.

15. Cornish S, Mehl-Madrona L. The role of vitamins and minerals in psychiatry. Integr Med Insights. 2008 Sep 24;3:33–42.

16. Alpert JE, Mischoulon D, Nierenberg AA, Fava M. Nutrition and depression: focus on folate. Nutrition. 2000 Jul;16(7-8):544–6.

17. Xu XJ, Jiang GS. Niacin-respondent subset of schizophrenia – a therapeutic review. Eur Rev Med Pharmacol Sci. 2015;19(6):988–97.

18. U S Department of Health and Human Services. The international classification of diseases. World Health Organization; 1969.

19. 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.

20. Rogers W, Lang F, Gay C. MetaMap Data File Builder. US National Library of Medicine. 2012.

21. CDC. International classification of diseases, ninth revision, clinical modification (ICD-9-CM). URL: http://www.cdc.gov/nchs/about/otheract/ict9/abticd9.htm [accessed 2004 Dec 16]. 2013;

22. HCUP-US Tools & Software Page [Internet]. [cited 2017 Sep 12]. Available from: https://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp

23. MIMIC [Internet]. [cited 2017 Sep 12]. Available from: https://mimic.physionet.org/about/mimic/

24. Fact Sheet MEDLINE®. 2004 Nov 17 [cited 2017 Sep 12]; Available from: https://www.nlm.nih.gov/pubs/factsheets/medline.html

25. National Library of Medicine - National Institutes of Health. 1993 Oct 10 [cited 2017 Sep 12]; Available from: https://www.nlm.nih.gov/

26. Fact Sheet Medical Subject Headings (MeSH®). 1999 Sep 1 [cited 2017 Sep 12]; Available from:
https://www.nlm.nih.gov/pubs/factsheets/mesh.html

27. Office of Dietary Supplements - PubMed Dietary Supplement Subset [Internet]. [cited 2017 Sep 12]. Available from: https://ods.od.nih.gov/Research/PubMed_Dietary_Supplement_Subset.aspx

28. Office of Dietary Supplements (ODS) [Internet]. [cited 2017 Sep 12]. Available from: https://ods.od.nih.gov/

29. 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

30. Dietary Supplement Label Database (DSDL) [Internet]. [cited 2017 Jul 18]. Available from: https://dsld.nlm.nih.gov/dsld/index.jsp

31. 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/substanceregistrationssystem-uniqueingredientidentifierunii/

32. 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

33. Welcome to the Natural Medicines Research Collaboration [Internet]. [cited 2017 Jul 18]. Available from: https://naturalmedicines.therapeuticresearch.com/

34. Latest Medical News, Clinical Trials, Guidelines – Today on Medscape [Internet]. [cited 2017 Jul 18]. Available from: http://www.medscape.com/

35. National Drug File - Reference Terminology Source Information. 2016 Jan 14 [cited 2017 Sep 12]; Available from: https://www.nlm.nih.gov/research/umls/sourcereleasedocs/current/NDFRT/

36. Liu S, Ma W, Moore R, Ganesan V, Nelson S. RxNorm: prescription for electronic drug information exchange. IT Prof. 2005;7(5):17–23.

37. Medical Subject Headings - Home Page. 1999 Sep 1 [cited 2017 Jul 18]; Available from: https://www.nlm.nih.gov/mesh/

38. Dervaux A, Laqueille X. [Thiamine (vitamin B1) treatment in patients with alcohol dependence]. Presse Med. 2017 Mar;46(2 Pt 1):165–71.

39. Gao L, Liu X, Yu L, Wu J, Xu M, Liu Y. Folic acid exerts antidepressant effects by upregulating brain-derived neurotrophic factor and glutamate receptor 1 expression in brain. Neuroreport [Internet]. 2017 Sep 12; Available from: http://dx.doi.org/10.1097/WNR.0000000000000887

40. Sharma V, Holmes JH, Sarkar IN. Identifying Complementary and Alternative Medicine Usage Information from Internet Resources. A Systematic Review. Methods Inf Med. 2016 Aug 5;55(4):322–32.

41. Wang Y, Adam TJ, Zhang R. Term Coverage of Dietary Supplements Ingredients in Product Labels. AMIA Annu Symp Proc. 2016;2016:2053–61.

42. 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.

43. 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.