



Published in final edited form as:

*Artif Intell Med.* 2016 March ; 68: 47–57. doi:10.1016/j.artmed.2016.02.003.

## A method for the development of disease-specific reference standards vocabularies from textual biomedical literature resources

Liqin Wang<sup>a,b,\*</sup>, Bruce E. Bray<sup>a,c</sup>, Jianlin Shi<sup>a</sup>, Guilherme Del Fiol<sup>a</sup>, and Peter J. Haug<sup>a,b</sup>

<sup>a</sup>Department of Biomedical Informatics, University of Utah, 421 Wakara Way, Salt Lake City, UT 84108, USA

<sup>b</sup>Homer Warner Research Center, Intermountain Healthcare, 5121 South Cottonwood Street, Murray, UT 84107, USA

<sup>c</sup>Department of Internal Medicine, University of Utah, 30 North 1900 East, Salt Lake City, UT 84132, USA

### Abstract

**Objective**—Disease-specific vocabularies are fundamental to many knowledge-based intelligent systems and applications like text annotation, cohort selection, disease diagnostic modeling, and therapy recommendation. Reference standards are critical in the development and validation of automated methods for disease-specific vocabularies. The goal of the present study is to design and test a generalizable method for the development of vocabulary reference standards from expert-curated, disease-specific biomedical literature resources.

**Methods**—We formed disease-specific corpora from literature resources like textbooks, evidence-based synthesized online sources, clinical practice guidelines, and journal articles. Medical experts annotated and adjudicated disease-specific terms in four classes (i.e., *causes or risk factors, signs or symptoms, diagnostic tests or results, and treatment*). Annotations were mapped to UMLS concepts. We assessed source variation, the contribution of each source to build disease-specific vocabularies, the saturation of the vocabularies with respect to the number of used sources, and the generalizability of the method with different diseases.

**Results**—The study resulted in 2588 string-unique annotations for heart failure in four classes, and 193 and 425 respectively for pulmonary embolism and rheumatoid arthritis in treatment class. Approximately 80% of the annotations were mapped to UMLS concepts. The agreement among heart failure sources ranged between 0.28 and 0.46. The contribution of these sources to the final vocabulary ranged between 18% and 49%. With the sources explored, the heart failure vocabulary reached near saturation in all four classes with the inclusion of minimal six sources (or between four to seven sources if only counting terms occurred in two or more sources). It took fewer sources to reach near saturation for the other two diseases in terms of the treatment class.

\*Corresponding author at: Department of Biomedical Informatics, University of Utah, 421 Wakara Way, Salt Lake City, UT 84108, USA.liqin.wang@utah.edu (L. Wang).

Appendix A. Supplementary data: Supplementary data associated with this article can be found, in the online version, at [http://dx.doi.org/10.1016/\\_j.artmed.2016.02.003](http://dx.doi.org/10.1016/_j.artmed.2016.02.003).

**Conclusions**—We developed a method for the development of disease-specific reference vocabularies. Expert-curated biomedical literature resources are substantial for acquiring disease-specific medical knowledge. It is feasible to reach near saturation in a disease-specific vocabulary using a relatively small number of literature sources.

### Keywords

Knowledge extraction; Reference standards; Annotation; Saturation; Disease-specific ontology; Heart failure

---

## 1. Introduction

Disease-specific ontologies are knowledge bases intended to structure and represent disease-relevant information including disease etiology, diagnosis, treatment and prognosis. The availability of these ontologies could facilitate cross-disciplinary exchange and sharing of domain-specific knowledge. Disease-specific ontologies are also essential in supporting a variety of domain-specific computer applications, such as natural language processing, cohort selection, and clinical decision support [1,2]. For example, Haug et al. initiated a pneumonia-specific ontology, which supported the development of a clinical diagnostic modeling system [3]. Malhotra et al. constructed an Alzheimer's disease ontology and applied it to text mining on electronic health records [4].

However, the lack of comprehensive disease-specific ontologies hinders the development of such applications. BioPortal [5], an open repository of biomedical ontologies, currently hosts up to 467 ontologies in various domains. However, among those ontologies less than 1% are disease-specific. Therefore, methods are needed to help develop disease-specific ontologies that can be made available to the community. A long-term goal of our research is to enable a platform that supports large-scale development of such ontologies.

Creating disease-specific ontologies is still a labor-intensive process. One of the main challenges is the knowledge acquisition, i.e., comprehensively ascertaining domain-specific concepts and relationships in the ontologies [6,7]. In knowledge engineering, domain experts are often used as the sources for acquiring medical knowledge. However, they are scarce and expensive. Another challenge is that while existing large terminologies, such as Disease Ontology [8] and Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) [9], can be used as sources of concepts for disease ontologies, the relationships between the concepts are primarily hierarchical, with little non-hierarchical relations between diseases and their signs and symptoms, diagnostic procedures, and treatments. Therefore, it is not feasible to extract a comprehensive set of disease-related relationships from those terminologies.

A promising alternative to address disease-ontology development challenges is to learn ontologies from textual data [7,10]. The learning can be separated into multiple levels: learning terms, synonyms, concepts, relations, axioms and rules [7,10]. At the term level, for instance, Riloff proposed a corpus-based approach for building domain-specific semantic lexicons [11]. At relationship level, Sanchez and Moreno studied methods that learn non-taxonomic relationships from web documents [12]. Particularly for developing disease-

specific ontologies, the learning is primarily focused on using narrative text sources, such as the biomedical literature, to automatically identify disease-relevant concepts and relations. The relationships include the taxonomy backbone (i.e., is-a relations) and non-hierarchical relations (e.g., treats, causes). Most hierarchical relations between biomedical concepts are well represented in large domain ontologies and terminologies, such as SNOMED CT and the Unified Medical Language System (UMLS). However, important gaps still exist in regards to non-hierarchical relations. Learning these relations is an active subject of research interest [13–17].

The goal of the present study is to design and test a generalizable method for the development of vocabulary reference standards from expert-curated, domain-specific documents, such as textbooks, and clinical guidelines. The vocabularies and analyses established will be used to help the development and testing of automated disease-specific knowledge acquisition algorithms.

In the process of developing reference vocabularies, the number and types of sources that are needed to maximize the number of concepts retrieved are unknown. One source is unlikely to provide all concepts and relations about a disease and it is not feasible to manually extract concepts from all literature sources available. Therefore, in present study, we investigate the number of sources that are needed to obtain saturation for a disease-specific vocabulary. We assessed the feasibility of acquiring disease-specific concepts and relationships in the classes of *causes and risk factors*, *sign and symptoms*, *diagnostic tests and results*, and *treatments* by manually annotating terms from a representative and diverse set of popular knowledge sources in cardiology. Last, we then tested the generalizability of our methods with two additional diseases in *treatment* class.

## 2. Methods

In the present study, a disease-specific vocabulary is understood to be a list of concepts that are semantically related to a disease or syndrome. We focused on gathering disease-related concepts into a collection rather than identifying their taxonomic structure [18]. The framework for acquiring disease-specific vocabulary is displayed in Fig. 1. This is an iterative process with the goal of reaching near saturation which in this study is defined as finding <5% new concepts with the introduction of a new resource. We first formed a corpus with a collection of textual biomedical literature documents on the topical disease. Then, we initiated the iterative process by selecting one source from the corpus. The following step is annotation and adjudication, where the documents were annotated using eHOST, an open source annotation tool [15], by medical experts based on an annotation guideline. Any conflicted annotations were adjudicated by consensus between experts. Annotations were then mapped to equivalent UMLS concepts or, if these were unavailable, assigned to local codes. Subsequently, we annotated each additional source, assessing the saturation of the vocabulary after each new source was included. The iteration ended when the vocabulary reached near saturation. More details are provided in the following sections.

## 2.1. Selection and preparation of knowledge sources

There are a large number of textual resources available that provide disease-specific medical knowledge, such as textbooks, online evidence-based documents, journal articles, medical records, etc. When choosing source documents, we prefer those that are expert-curated, knowledge-dense, and evidence-based. As a starting point, the following types of knowledge sources were chosen: regularly updated textbooks, evidence-based synthesized online sources, clinical practice guidelines, and disease-specific journal articles. Medical textbooks provide a comprehensive and general overview of select diseases from their diagnosis to treatment. Two decades ago, Curley investigated physician's preferences for acquiring medical knowledge and found that among the many resources textbook and journals were most frequently used [19]. Nowadays, they are still among the conventional sources used by medical students to obtain medical knowledge. Another kind of source, clinical practice guidelines, have been made available for many medical domains and are used to “assist practitioner and patient decisions about appropriate health care for specific clinical circumstances. [20] They typically focus on therapeutic guidance. Due to the high demand for speed in answering daily clinical questions, online, point-of-care, evidence-based products have become available and have gained wide acceptance from healthcare professionals. Examples of these products include Clinical Evidence, DynaMed, InfoRetriever, PDxMD and UpToDate [21].

We consulted domain experts to choose one or two examples of each type of source mentioned above. For heart failure, a total of seven source documents were chosen (see Table 1): Braunwald's heart disease (Braunwald) [22], Harrison's principle of internal medicine (Harrison's) [23], UpToDate®, DynaMed™, American College of Cardiology Foundation (ACCF)/American Heart Association (AHA) guidelines for the management of heart failure (ACC guideline) [24,25], European Society of Cardiology (ESC) guidelines for the diagnosis and treatment of acute and chronic heart failure (ESC guideline) [26], and ACC/AHA key data elements and definitions (ACC key data elements) [27]. The publication date, popularity (number of citations), and the sections of each of the sources used in this study are further specified in Table 1. In the same way, we formed corpus for two other conditions. For pulmonary embolism, the initial corpus includes a clinical guideline [28], chapters from a textbook (Braunwald) [22], four articles from UpToDate®, and one document from DynaMed™. For rheumatoid arthritis, the corpus includes four articles from UpToDate®, one document from DynaMed™, and one chapter from Harrison's.

## 2.2. Annotation scheme

As we examined those expert-curated, disease-specific documents, we found that their content covers different aspects of a disease including etiology, diagnosis, treatment/prevention, and prognosis. To maintain feasibility for the annotation tasks, we provided a more granular classification, and restricted the annotations to four class types: *causes or risk factors*, *signs or symptoms*, *diagnostic tests or results*, and *treatment*; other classes were excluded from this annotation task, such as comorbidities and complications. Table 2 gives more detail about the definitions and examples for these annotation classes.

### 2.3. Annotation and adjudication

Annotation is a process to identify salient terms from a collection of narrative documents and assign them to a proper class, while adjudication is a process to resolve conflicting results between annotators. We used eHOST [32] for both processes. Besides annotation, eHOST supports dictionary export, pre-annotation, and measurement of inter-annotator agreement (IAA).

Initially, we developed an annotation guideline (available in the online supplement) with rules and classes. Before the actual annotation task, annotators went through a training process, and annotated a sample collection of documents. The degree of IAA between the two annotators was calculated using F-measure where we treated one annotator as the subject and the second annotator's results as if they were a gold standard [33]. Text annotation can be seen as an information retrieval task therefore normal Kappa statistics cannot be calculated without a negative case count. Annotators iteratively annotated sample documents until the IAA score reached substantial agreement (IAA between 0.6 and 0.8) [34]. Then, annotators began working on the same set of documents from the corpora. Disagreements were resolved through a consensus process between two annotators. However, for the diseases in which the annotators reached almost perfect agreement (IAA greater than 0.8) on the first document, we proceeded to the subsequent documents with only one annotator per document.

We also used pre-annotation to improve the annotation quality and efficiency as suggested in previous studies [35,36]. Annotators were firstly assigned a small set of documents to annotate. After adjudication, we extracted a list of terms as a dictionary from these documents. Next, the subsequent documents were automatically pre-annotated with the dictionary compiled from previously adjudicated annotations. With these pre-annotated documents, annotators could modify or delete pre-annotations, or add missed occurrences of terms.

### 2.4. Mapping annotations to UMLS concepts

The annotation texts contain various lexical variations such as abbreviations/acronyms (e.g., “EF” for “*ejection fraction*”), synonyms (e.g., “*alcohol consumption*” and “*alcohol intake*”), compound terms (e.g., “*cardiac catheterization and revascularization*”, “*coronary or peripheral vascular disease*”), and modifiers (e.g., “*daily serum electrolytes*”), which make it difficult to compare the annotations among the sources. To address this issue, we sought to map all the annotations from these source documents to standard terminologies, including SNOMED CT, Logical Observation Identifiers Names and Codes (LOINC), RxNorm and Medical Subject Headings (MeSH). This was expected to facilitate an analysis of the vocabulary sets obtained from different sources and used to form a final vocabulary. The terms that did not correspond to entries in standard terminologies may, in the future, be used to enhance existing terminologies. The UMLS Metathesaurus is the largest thesaurus in the biomedical domain. It has integrated hundreds of source terminologies, and provides cross-mapping to different source terminologies. We mapped the annotations to UMLS concepts while restricting the source terminologies to the four mentioned above. The 2014AB version was used in this study.

Concept mapping can be a subjective task. Lexical variations increase the complexity of the mapping. For some terms it may not even be possible to find mappings from the UMLS Metathesaurus. In order to reduce the subjectivity of the mapping and to make the mapping process more reliable and reproducible, we set up several mapping rules (see Table 3). For example, we restricted the semantic types for mapped concepts to the *treatment* and *diagnostic tests or results* classes. For terms with modifiers (e.g., daily serum electrolytes), we used a post-coordinated mapping approach that we removed the modifier (e.g., daily) and assigned the core term (e.g., serum electrolytes) with equivalent UMLS concept unique identifier (CUI). Compound terms (e.g., atrial and ventricular arrhythmias) were extended and mapped to multiple terms.

The mappings were automatically processed by MetaMap [37] and followed by manual verification and selection of concepts. Mapping rules were applied for some special cases (see examples in Table 3). The annotations that were not mappable to any targeted terminologies in any form were temporarily assigned local codes in order to support the analysis (e.g., atrial fibrillation surgery → atrial fibrillation surgery - Local Code: T0000001). The manual verification and selection were mainly done by one investigator (LW), with some assistance from a cardiologist (BEB). We also tested the agreement of the mappings by comparing the mapping results against mappings from another individual (JS) on 237 randomly sampled terms. The agreement of the mappings between these two individuals in terms of F-measure [33] was 0.84.

## 2.5. Saturation assessment

A natural process to build a vocabulary from a corpus (or knowledge sources) is to add all the acquired vocabulary from it at once. However, in order to answer our research question, we analyzed the accumulation process of the acquired vocabulary where we take the sources one by one, and determine whether, at a certain point, the vocabulary reaches some level of saturation. The accumulation rate of the vocabulary is calculated by the ratio of the number of new concepts from the last included source divided by the total number of concepts from all included sources (see Formula (1)). We considered a concept to be new when its code (either a UMLS CUI or temporary assigned code) did not appear in the vocabulary from previous documents. We chose 5% as an arbitrarily threshold to determine near saturation, i.e., a small number of new concepts are added by including new sources.

$$\text{Accumulation rate} = \frac{\text{the number of new concepts } (S_n)}{\text{the number of all concepts } \left( \sum_n^1 S_i \right)} \quad (1)$$

where  $S_i$  is the  $i$ th entered source.

The accumulation rate is the key factor used in this study to determine whether a vocabulary has reached near saturation or not. However, the rates can be affected by the order in which sources are included. Supposing that most of the concepts from one entered source have been presented in the existing vocabulary, the accumulation rate could drop significantly; however the rate could increase again if a subsequent source has substantially different



vocabulary from others. To adjust for this situation, we determine the order of the sources in a given corpus by maximizing or minimizing the accumulation rate at each step. In another words, at each accumulation step the next incoming source is determined by selecting the source that can achieve the highest or lowest accumulation rate among the remaining sources. This method helps determine the lower bound and/or upper bound of the number of sources for reaching near saturation for a disease-specific vocabulary with the corpus explored. When the order of the sources changes, the number of sources for reaching near saturation falls in that range.

Although all the chosen sources are expert-curated and the annotation was done by medical experts, it does not guarantee all acquired relations are clinically-valid. For example, annotating conclusion sentences from a single clinical trial study could bring relations that are not clinically-valid into the vocabulary. We believe that terms that only occur in one source should be treated as less valid than those that occur in multiple sources. To address this issue, we include an analysis based on the concepts that appear in two or more sources which we called core concepts.

The saturation measurement is source-dependent. If the sources consistently overlap, then a smaller number of input sources are needed in order to reach a saturated status. Therefore, we use F-measure to measure the agreement of the chosen sources to indicate the degree of the source variety.

### 3. Results

For heart failure, two annotators reached substantial agreement on the first set of documents and therefore both were involved in the annotation and adjudication on the seven source documents. The annotation process resulted in 2588 string-unique annotations, which were mapped to 1648 concepts. The majority of these annotations ( $N=2109$ , 81%) was mapped to 1232 UMLS concepts with the sources restricted to SNOMED CT, LOINC, RxNorm, and MeSH. The remainder were mapped to 416 local terms. For the other two conditions, two annotators reached perfect agreement in the first documents, and therefore, the subsequent documents were processed with only one annotator per document. For pulmonary embolism, we retrieved 193 string-unique annotations related to *treatment*, which were mapped to 96 concepts (83% were mapped to UMLS concepts). Similarly, for rheumatoid arthritis, we obtained 425 string-unique annotations, which were mapped to 279 concepts (83% were mapped to UMLS concepts). Independent of the diseases and class types, we obtained 3142 string-unique annotations, which mapped to 2049 concepts. On average, each concept had 1.5 corresponding terms or synonyms.

#### 3.1. Concepts distributed by sources and classes

Fig. 2 shows the number of unique concepts per class across seven knowledge sources related to heart failure. Among the seven sources, Braunwald had the largest number of concepts in the classes of *signs or symptoms*, *diagnostic tests or results*, and *treatment*, while the ACC guideline had the largest amount of concepts in the class of *causes or risk factors*. Fig. 2 also shows the number of acquired concepts along with the size of each document based on word count. Harrison's is the smallest corpus, containing 10,560 words, and from

which we obtained 298 concepts, while Braunwald is the largest corpus (98,306 words) and from which we obtained 810 concepts. However, the size of the documents was not proportional to the number of obtained concepts. For example, UpToDate is almost twice as large as ACC key data elements and ESC guideline (20,661 vs. 9344 and 10,452 words); however, it provided a slightly smaller vocabulary of concepts (380 vs. 386 and 481 concepts).

Table 4 shows the contribution of each knowledge source to individual classes as well as the final heart failure vocabulary. The contribution of the seven knowledge sources ranged between 13–63% for the individual classes and between 18–49% for the final vocabulary. Among the seven sources, ACC guideline had the best contribution to the class of *causes or risk factors*, while Braunwald had best contribution to the other three classes. Some concepts were assigned to multiple classes, which explains why the sum total of the four classes is not equal to the total number of the vocabulary. For example, "hypertension" was assigned to the class of *causes or risk factors* and *diagnostic tests or results in a different context*.

### 3.2. Distribution of concept occurrence

Fig. 3 shows the log-log-scale distribution of the number of heart-failure concepts by concept frequencies. 747 concepts (45% of the total) were annotated only once in the entire corpus, and 247 concepts (15% of the total) were occurred twice, however, some concepts were annotated much frequently, such as one concept was annotated 471 times. Overall, the log-log-scale distribution appears approximately linear. Table 5 lists a small set of concepts that frequently occurred in the corpus, where the frequency of concept is measured by the occurrence of the corresponding annotations over the corpus.

### 3.3. Variety of the sources

Table 6 shows the agreement of concepts among the seven sources. The overall agreement between source pairs ranged from 0.28 to 0.46. From Table 6, ACC key data element consistently had lower agreement with other sources. Sources from the same category (e.g., textbook), such as Braunwald and Harrison's, did not show a stronger agreement than sources from different categories.

### 3.4. Concept accumulation by classes

With the explored corpus, Fig. 4A shows the minimum accumulation of heart failure concepts with the inclusion of additional sources. The accumulation curves of all four classes appear approximately linear. The number of concepts per class ranged from 170 concepts in *signs or symptoms* to 445 concepts in *diagnostic tests or results*. Fig. 4B shows the maximum accumulation of heart failure concepts. The accumulation curves of all four classes increase quickly with the inclusion of the first set of sources, but reach a plateau around the inclusion of the sixth source.

Fig. 4C and D show the accumulation rates with the inclusion of new sources (percent of identified concepts that are new) with two different orders of sources, where the 4C achieved the minimal accumulation rates and 4D achieved the maximum accumulation rates. The curves in Fig. 4C decline at the beginning, however, stay flatten around 20% accumulation



rate since the inclusion of the fourth sources, and even increase slightly with the introduction of the sixth or seventh sources. We found that Braunwald is the last source in the minimum accumulation, which contained almost half amount of the concepts in the final vocabulary. The curves in Fig. 4D decline steeply with the inclusion of the second and third sources and are flatten after the inclusion of the fourth source. Although the number of concepts per class is different (from Fig. 4A), the accumulation rates are similar for all classes. After the sixth source, very few concepts are added to the vocabularies.

Fig. 5A and B show the minimum and maximum accumulation of the number of core concepts (i.e., concepts occur in two or more sources) respectively to the inclusion of additional sources. Fig. 5C and D are the corresponding analysis of the accumulation rates. Based on these two figures, the heart failure vocabulary reaches near saturation by using between four to seven sources regardless of the order the sources. Comparing the minimal accumulation rates between 4C and 5C, the accumulation rates in 5C decline faster with all four classes, and reach the 5% threshold after adding the seventh source, while in 4C, the accumulation rates never reach the 5% threshold. The accumulation rates of all four classes in 5D decline slightly faster than the rates in 4D, and reach the 5% threshold after adding the fourth source comparing sixth source in 4D. This confirms that concepts that are used more often can be gathered with fewer knowledge sources. Table 7 provides sample terms that occurred in one and multiple sources for each class.

### 3.5. Concept accumulation by conditions

A similar decline in the benefit of reviewing additional sources was seen in the other two diseases studied. Fig. 6A shows the maximum accumulation of treatment concepts along with increase of corpus size. For different diseases, the number of retrieved treatment concepts varies. From Fig. 6B, we found that the accumulation rates of heart failure decreased to the 5% threshold after adding the sixth source, while it only took three sources for rheumatoid arthritis to reach the threshold and four sources for pulmonary embolism.

Fig. 7A and B show the minimum and maximum accumulation rates of core *treatment* concepts per condition along with the addition of new sources. Based on these two figures, heart-failure *treatment* vocabulary of core concepts reached near saturation by using between three and seven sources; rheumatoid arthritis used between two and three sources; and pulmonary embolism used between two and four sources. All three conditions reached near saturation with a relatively small number of sources regardless of the order of sources.

## 4. Discussion

In this study, we assessed the feasibility of obtaining a near-saturated disease-specific vocabulary using a diverse set of expert-curated textual knowledge sources. We also estimated the number of sources needed to reach near saturation with four disease concept classes (i.e., *causes or risk factors*, *sign or symptoms*, *diagnostic tests or results*, and *treatment*) with heart failure. We tested the generalizability of the method with two other conditions, i.e., pulmonary embolism, and rheumatoid arthritis.

From the study, we found that regardless of the difference of the total number of acquired concepts, the vocabularies of four concept classes reached near saturation at similar pace. For the four concept classes explored on the heart failure condition, the vocabularies took six sources to reach near saturation with the best order of sources. However, when order changes, they may require more sources to reach near saturation. When considering only the core concepts, three conditions all achieved near saturation with much fewer sources regardless of the order of sources. Overall, the results support the conclusion that it is feasible to obtain near-saturated reference standards for disease-specific vocabularies of core concepts using a relatively small number of knowledge sources. If choosing preferentially the sources with big contribution, the vocabularies of all concepts can also achieve near saturation with a relatively small number of sources.

The main contribution of the study lies in two aspects. First, the findings of this study are important for the development of disease-specific reference vocabularies. Developing reference standards usually involves substantial manual knowledge acquisition effort. The results of our study provide an estimate for the optimal number of text sources that can be used to find a balance between cost and saturation. Second, the method proposed in the present study provides an underlying approach for the development of disease-specific vocabularies. This includes the selection of sources, the guideline for annotation, the rules for mapping, and saturation assessment. Our approach is designed to be efficient as the method is able to determine a stopping point where a vocabulary has reached near saturation. Since the method had similar results in different conditions and concept classes, it is expected that the method and results will generalize to other conditions. We intend to use this approach to develop other disease-specific vocabularies and use them as reference standards for the development and testing of automated methods to generate disease-specific concept vocabularies. Xu et al. provides an example of such an automated method that could benefit from our reference standards [17]. Their method used known disease-drug pairs to learn patterns from biomedical text, enabling an automated, large-scale extraction of drug-disease treatment pairs. The reference standards can also be used to assess the performance of the automated systems by comparing the automated generated concepts to those manually extracted disease-specific vocabularies.

Seven authoritative sources that are frequently used by clinicians were chosen for heart failure. No previous study has investigated these sources regarding their contribution to a specific topic and the overlap among the sources. From our results, the agreement scores of the concepts among the seven sources ranged from 0.28 to 0.46, which indicates that these sources did not strongly overlap. The study results also suggest that specialized textbooks (e.g., Braunwald for heart failure) should be used as a starting point for building domain-specific vocabularies, as they appear to provide the broadest contribution to the vocabulary compared to other sources. However, the results discourage using textbooks as the only source, as even the lengthiest textbook (Braunwald) only captured half of the domain concepts represented in the final vocabulary. The use of multiple and diverse sources is critical to construct a comprehensive vocabulary. When applying our method to other conditions, we recommend that the most optimal approach is to start the annotation with comprehensive textbooks on topical disease, followed by relevant and most-updated clinical

practice guidelines, and then topic summary articles from evidence-based synthesized online resources such as UpToDate.

The annotation process was sometimes complex and subjective. For example, for heart failure, it was difficult to discriminate between the treatment of its symptoms and treatments directed toward the causes and risk factors (e.g., hypertension, diabetes mellitus). A large portion of each source document was dedicated to discussing the treatment of the causes and risk factors of heart failure. Discriminating among these relationships is necessary to correctly associate diseases with their concepts.

Mapping annotations to standard terminologies or local terms is an essential step for saturation assessment. Without mapping, it may require to annotate a larger corpus to reach near saturation, however, which may only lead to more terms with all kind of varieties. After the mapping, we identify many terms ( $N=364$ ) that were not available in standard terminologies. These unmapped concepts could be used to enhance existing standard vocabularies. Another interesting finding is that almost half amount of concepts annotated only once in the corpus (see Fig. 3). Based on a manual review on the concepts with different occurrences, concepts that occurred only once or came from a single source (see Table 7) show less clinical relevance to the topic condition. In order to build a disease ontology with strong evidences, we may exclude those concepts. Besides, this distribution of concept occurrence over the corpus (see Fig. 3) almost follows a Zipf distribution. This could be possibly used for ranking strength of the relations of the concepts to the disease.

Our study has a few limitations. First, although the final vocabulary reached near saturation, we believe many disease-specific concepts are still missing. For example, “Angiotensin-convertingenzyme inhibitor” (ACE Inhibitor) – a class of drugs for treating heart failure, is present in the acquired vocabulary. But not all the drugs under this class were explicitly presented in the source documents. When experts mentioned the ACE inhibitor in those source documents, they were probably referring to the entire class of the drugs. These kinds of missing concepts could be inferred from relationships in existing standard terminologies, such as RxNorm and SNOMED CT. Second, the actual coverage of the final vocabulary is unknown. Assessing the actual coverage of a vocabulary is difficult because perfect reference vocabularies are not available. Third, the near saturation was reached with an optimal order of the sources. However, when the order of sources changed, the vocabulary may not be saturated. The upper bound of the number of sources for near saturation was not detected with a small number of sources explored in this study. However, for the vocabulary of core concepts is sufficient to reach near saturation with a relatively small number of sources regardless of the order of sources.

## 5. Conclusions

We provided an underlying approach for the development of disease-specific reference vocabularies focused on the concept classes of *causes and risk factors*, *signs and symptoms*, *diagnostic tests and results*, and *treatment*. Our findings show that expert-curated sources, such as textbooks, clinical guidelines, evidence-based summaries, and journal articles, are substantial sources for disease-specific medical knowledge. Their contribution to the

vocabulary varies substantially for a specified condition. While the numbers of sources for reaching near saturation can vary modestly for different conditions, a relatively small number of text sources are sufficient to obtain a near-saturated vocabulary of sound disease-specific concepts. In the future we intend to develop automated techniques to extract disease-specific vocabularies from large corpora. The reference standards developed in the present study will be used to assess the performance of the automated vocabulary extraction system.

### Data availability

The extracted heart-failure-specific vocabulary, comprising 1648 concepts in the aspect of causes and risk factors, signs or symptoms, diagnostic tests or results, and treatment, is available at <http://purl.bioontology.org/ontology/HFO>. The rheumatoid arthritis vocabulary contains 279 concepts in the aspect of treatment and is available at: <http://bioportal.bioontology.org/ontologies/RAO>, while pulmonary embolism vocabulary with 96 concepts is available at: <http://bioportal.bioontology.org/ontologies/PE>. The hierarchical relationships in these vocabulary are obtained from the UMLS.

### Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

### Acknowledgments

The authors thank Robert Hausam in the initial set up of annotation training, and Philip Brewster for the language editing. This work is supported in part by Grant LM010482 from the National Library of Medicine.

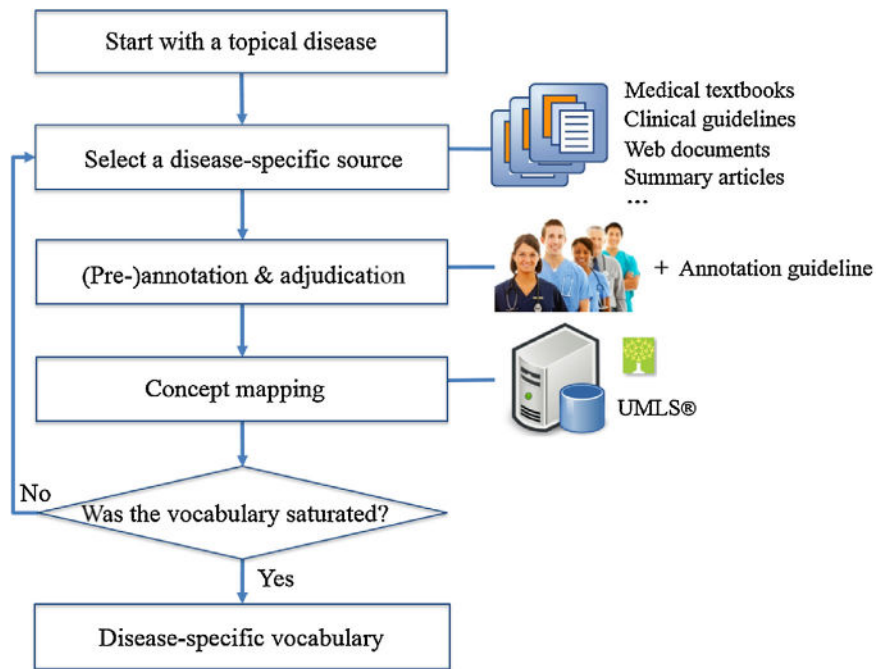
### References

1. Bodenreider O. Biomedical ontologies in action: role in knowledge management, data integration and decision support. *IMIA Yearb Med Inform.* 2008; 47:67–79.
2. Rubin DL, Shah NH, Noy NF. Biomedical ontologies: a functional perspective. *Brief Bioinform.* 2008; 9:75–90. <http://dx.doi.org/10.1093/bib/bbm059>. [PubMed: 18077472]
3. Haug PJ, Ferraro JP, Holmen J, Wu X, Mynam K, Ebert M, et al. An ontology-driven, diagnostic modeling system. *J Am Med Inform Assoc.* 2013; 20:e102–10. <http://dx.doi.org/10.1136/amiajnl-2012-001376>. [PubMed: 23523876]
4. Malhotra A, Younesi E, Gündel M, Müller B, Heneka MT, Hofmann-Apitius M. ADO: a disease ontology representing the domain knowledge specific to Alzheimer's disease. *Alzheimer's Dement.* 2014; 10:238–46. <http://dx.doi.org/10.1016/j.jalz.2013.02.009>. [PubMed: 23830913]
5. Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res.* 2009; 37:W170–3. <http://dx.doi.org/10.1093/nar/gkp440>. [PubMed: 19483092]
6. Noy NF, McGuinness DL. *Ontology development 101: a guide to creating your first ontology.* Stanford Knowl Syst Lab Tech Rep KSL-01-05 Stanford Med Informatics Tech Rep SMI-2001-0880. 2001
7. Buitelaar, P., Cimiano, P. *Ontology learning from text: methods, evaluation and applications.* Amsterdam, The Netherlands: IOS Press; 2005. doi:10.1.1.70.3041
8. Schriml LM, Arze C, Nadendla S, Chang YWW, Mazaitis M, Felix V, et al. Disease Ontology: a backbone for disease semantic integration. *Nucleic Acids Res.* 2012; 40:D940–6. <http://dx.doi.org/10.1093/nar/gkr972>. [PubMed: 22080554]

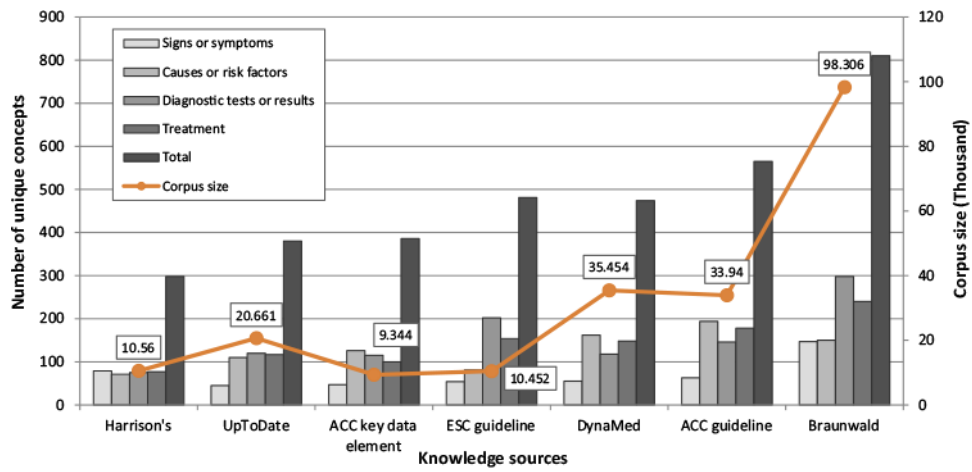
9. Stearns MQ, Price C, Spackman KA, Wang AY, Authority I, Cross A. SNOMED Clinical Terms: overview of the development process and project status. *AMIA Annu Symp Proc.* 2001:662–6.
10. *Ontology learning and population from text: algorithms, evaluation and applications.* Secaucus, NJ, USA: Springer-Verlag New York, Inc.; 2006. <http://dx.doi.org/10.1007/978-0-387-39252-3>
11. Riloff E, Shepherd J. A corpus-based approach for building semantic lexicons. *Proc Second Conf Empir Methods Nat Lang Process.* 1997:117–24.
12. Sanchez D, Moreno A. Learning non-taxonomic relationships from web documents for domain ontology construction. *Data Know Eng.* 2008; 64:600–23.
13. 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; 15:87–98. <http://dx.doi.org/10.1197/jamia.M2401>. [PubMed: 17947625]
14. Wang X, Chused A, Elhadad N, Friedman C, Markatou M. Automated knowledge acquisition from clinical narrative reports. *AMIA Annu Symp Proc.* 2008:783–7. [PubMed: 18999156]
15. Wright A, Chen ES, Maloney FL. An automated technique for identifying associations between medications, laboratory results and problems. *J Biomed Inform.* 2010; 43:891–901. <http://dx.doi.org/10.1016/j.jbi.2010.09.009>. [PubMed: 20884377]
16. Xu R, Li L, Wang Q. dRiskKB: a large-scale disease-disease risk relationship knowledge base constructed from biomedical text. *BMC Bioinform.* 2014; 15:105. <http://dx.doi.org/10.1186/1471-2105-15-105>.
17. Xu R, Wang Q. Large-scale extraction of accurate drug-disease treatment pairs from biomedical literature for drug repurposing. *BMC Bioinform.* 2013; 14:181. <http://dx.doi.org/10.1186/1471-2105-14-181>.
18. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med.* 1998; 37:394–403. [PubMed: 9865037]
19. Curley SP, Connelly DP, Rich EC. Physicians' use of medical knowledge resources: preliminary theoretical framework and findings. *Med Decis Mak.* 1990; 10:231–41. <http://dx.doi.org/10.1177/0272989X9001000401>.
20. Institute of Medicine (US) Committee to Advise the Public Health Service on Clinical Practice Guidelines. *Clinical practice guidelines: directions for a new program.* Washington, DC: National Academies Press (US); 1990.
21. Alper BS. Practical evidence-based Internet resources. *Fam Pract Manage.* 2003; 10:49–52.
22. Bonow, RO., Mann, DL., Zipes, DP., Libby, P. *Braunwald's heart disease: a textbook of cardiovascular medicine.* 9th. Philadelphia, PA: W.B. Saunders Company; 2011.
23. Longo, D., Fauci, A., Kasper, D., Hauser, S., Jameson, J., Loscalzo, J. *Harrison's principles of internal medicine.* 18th. New York: McGraw-Hill; 2011.
24. Hunt SA, Abraham WT, Chin MH, Feldman AM, Francis GS, Ganiats TG, et al. 2009 focused update incorporated into the ACC/AHA 2005 guidelines for the diagnosis and management of heart failure in adults: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *Circulation.* 2009; 119:e391–479. <http://dx.doi.org/10.1161/circulationaha.109.192065>. [PubMed: 19324966]
25. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Drazner MH, et al. ACCF/AHA guideline for the management of heart failure: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol.* 2013; 62:e147–239. <http://dx.doi.org/10.1016/j.jacc.2013.05.019>. [PubMed: 23747642]
26. McMurray JJV, Adamopoulos S, Anker SD, Auricchio A, Bohm M, Dickstein K, et al. ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. *Eur Heart J.* 2012; 33:1787–847. <http://dx.doi.org/10.1093/eurheartj/ehs104>. [PubMed: 22611136]
27. Radford MJ, Arnold JMO, Bennett SJ, Cinquegrani MP, Cleland JGF, Havranek EP, et al. ACC/AHA key data elements and definitions for measuring the clinical management and outcomes of patients with chronic heart failure: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Data Standards (Writing Committee to

- Develop Heart Failure Clinical Data Standards). *Circulation*. 2005; 112:1888–916. <http://dx.doi.org/10.1161/CIRCULATIONAHA.105.170073>. [PubMed: 16162914]
28. Konstantinides SV, Torbicki A, Agnelli G, Danchin N, Fitzmaurice D, Galiè N, et al. 2014 ESC guidelines on the diagnosis and management of acute pulmonary embolism. *Eur Heart J*. 2014; 35:3033–69. <http://dx.doi.org/10.1093/eurheartj/ehu283.3069a-3069k>. [PubMed: 25173341]
  29. Colucci WS. Evaluation of the patient with heart failure or cardiomyopathy. UpToDate. 2013
  30. Colucci WS. Treatment of acute decompensated heart failure: components of therapy. UpToDate. 2013
  31. Colucci WS. Overview of the therapy of heart failure due to systolic dysfunction. UpToDate. 2013
  32. South, BR., Shen, S., Leng, J., Duvall, SL., Chapman, WW. Proc 2012 Work Biomed Nat Lang Process. 2012. A prototype tool set to support machine-assisted annotation; p. 130-9.
  33. Hripesak G, Rothschild AS. Agreement, the F-measure, and reliability in information retrieval. *J Am Med Inform Assoc*. 2005; 12:296–8. <http://dx.doi.org/10.1197/jamia.M1733>. [PubMed: 15684123]
  34. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977; 33:159–74. [PubMed: 843571]
  35. Fort K, Sagot B. Influence of pre-annotation on POS-tagged corpus development. Fourth ACL Linguist Annot Work. 2010:56–63.
  36. Lingren T, Deleger L, Molnar K, Zhai H, Meinzen-Derr J, Kaiser M, et al. Evaluating the impact of pre-annotation on annotation speed and potential bias: natural language processing gold standard development for clinical named entity recognition in clinical trial announcements. *J Am Med Inform Assoc*. 2014; 21:406–13. <http://dx.doi.org/10.1136/amiajnl-2013-001837>. [PubMed: 24001514]
  37. Aronson AR. Effective mapping of biomedical text to the UMLS Meta thesaurus: the MetaMap program. *AMIA Symp*. 2001:17–21. doi:D010001275. pii.

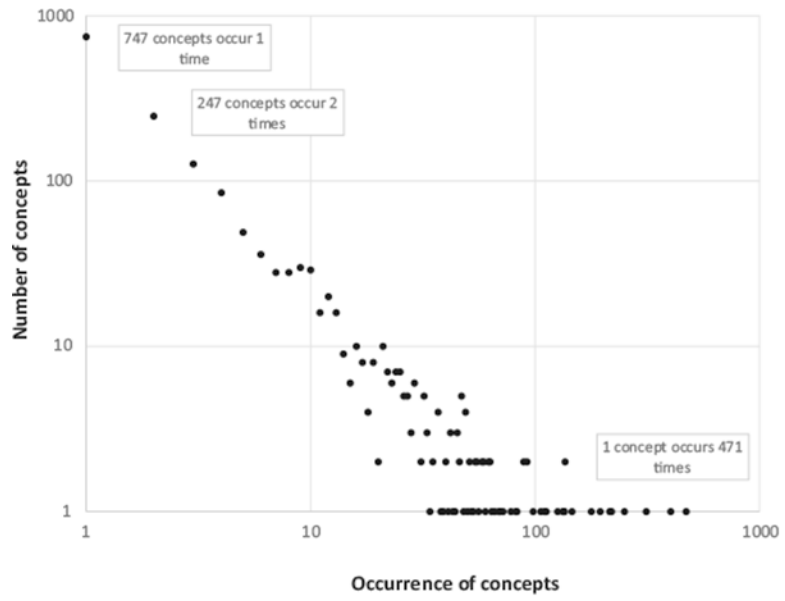




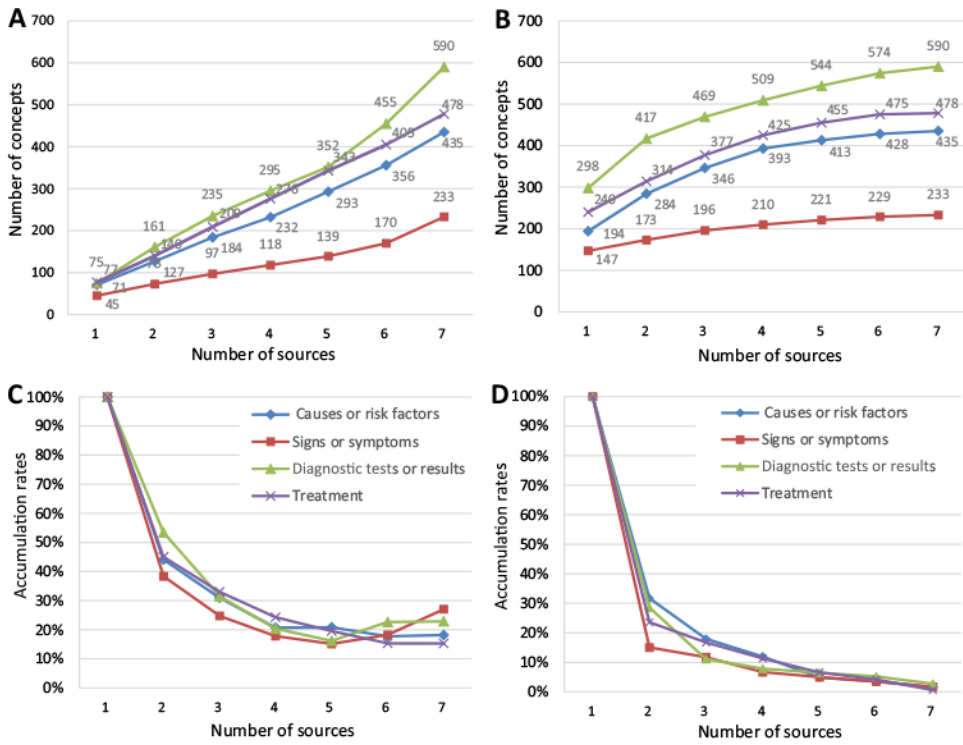
**Fig.1.** Workflow for building near-saturated, disease-specific reference vocabularies from biomedical literature resources.



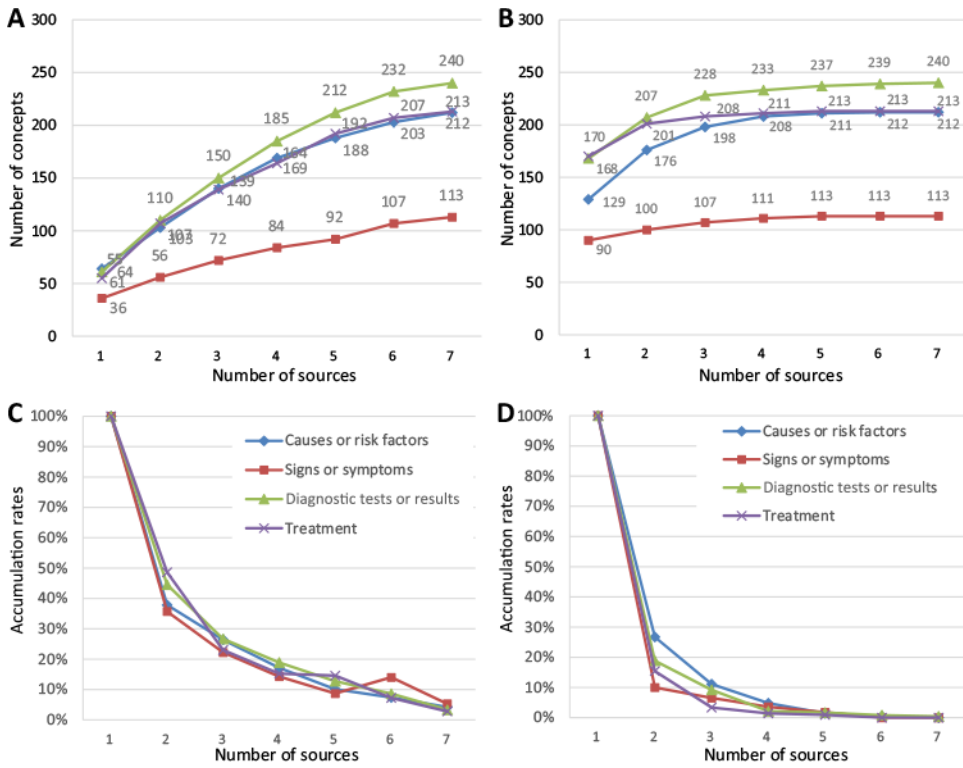
**Fig.2.** Distribution of heart failure concepts extracted from different knowledge sources for signs or symptoms, causes or risk factors, diagnostic tests or results, and treatment. The numbers in the rectangles represent the corpus size (word count) of each source



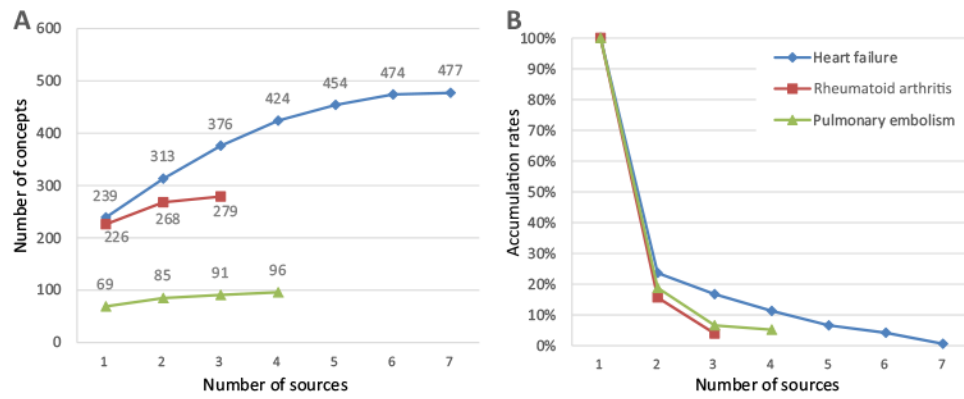
**Fig.3.** Log-log scale plot of the distribution of the number of heart failure concepts by concept occurrence.



**Fig.4.** (A) Number of heart failure concepts per class with the addition of new sources in minimum accumulation; (B) number of heart failure concepts per class with the addition of new sources in maximum accumulation; (C) minimum accumulation rates of heart failure concepts per class with the addition of new sources; (D) maximum accumulation rates of heart failure concepts per class with the addition of new sources.

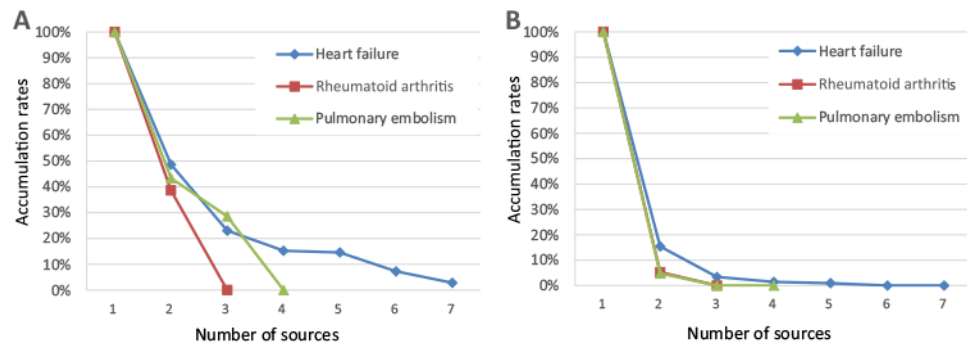


**Fig.5.** (A) Number of heart-failure core concepts (occurred in two or more sources) with the addition of new sources in minimum accumulation; (B) number of heart-failure core concepts with the addition of new sources in maximum accumulation; (C) minimum accumulation rates of heart-failure core concepts per class with the addition of new sources; (D) maximum accumulation rates of heart-failure core concepts per class with the addition of new sources.



**Fig.6.** (A) Number of *treatment* concepts per disease with the addition of new sources; (B) accumulation rates of *treatment* concepts per disease with the addition of new sources.





**Fig.7.**

(A) Minimum accumulation rates of core *treatment* concepts per disease with the addition of new sources; (B) maximum accumulation rates of core *treatment* concepts per disease with the addition of new sources

**Table 1**

Textual knowledge sources for extracting heart-failure-related concepts used to build a disease-specific vocabulary.

Sources	Types	Published/updated by	No. of citations	Included chapters/articles
Braunwald	Textbook	2011	4917	Chapter 26–34, excluded all the figures and references
Harrison's	Textbook	2011	9851	Chapter 234, exclude all the figures and references
UpToDate®	Evidence-based systematic online journal reviews	Dec 2013	N/A	The first three articles retrieved with the query of heart failure in adult [29-31]
DynaMed™	Evidence-based systematic online journal reviews	Feb 2014	N/A	The first document retrieved with the querying of "heart failure", include following sections: causes and risk factors, history and physical, diagnosis, treatment, and prevention and screening.
ACC guideline	Clinical practice guidelines	2009; 2013	1334; 535	For 2009 version, we include all the recommendations which are in bold, and all the tables. For 2013 version, all the content is included except the figures and references.
ESC guideline	Clinical practice guidelines	2012	1951	Included following chapters: 3–5, 7–10, 12–14.
ACC key data elements	Conclusive journal articles	2005	134	Section III. Heart failure clinical data standard elements and definitions

**Table 2**

Annotation scheme, definitions and examples.

Classes	Description	Examples
Causes or risk factors	Merges two overlapped subclasses: causes and risk factors, where causes refer to concepts or terms that can directly cause heart failure and risk factors are those factors associated with an increased risk of heart failure.	In industrialized countries, coronary artery disease (CAD) has become the predominant cause in men and women and is responsible for 60-75% of cause of heart failure. Hypertension contributes to the development of heart failure in 75% of patients.
Signs or symptoms	Groups medical signs and symptoms. In addition it includes the physical examination for which would usually result signs or symptoms.	The cardinal symptoms of heart failure are fatigue and shortness of breath. Nocturnal cough is a common manifestation of this process and a frequently overlooked symptom of heart failure.
Diagnostic tests or results	Includes phrases that describe procedures, panels, and measures that are done to a patient or a body fluid or sample in order to discover, rule out, or find more information about a medical problem.	A routine 12-lead ECG is recommended. A chest X-ray provides useful information about cardiac size and shape.
Treatment	Includes phrases that describe procedures, interventions, and substances given to a patient in an effort to resolve a medical problem.	Dietary restriction of sodium (2-3 g daily) is recommended in all patients with the clinical syndrome of heart failure and either preserved or depressed ejection fraction. Diuretics are the only pharmacologic agents that can adequately control fluid retention in advanced heart failure.

Table 3

Rules for mapping the annotations to UMLS concepts.

Mapping rules	Example
1. Map the terms to concepts that convey the term specific meaning within the context of the original sentence.	Sentence: "History of exposure to cardiotoxic substances through substance abuse: cocaine, amphetamine, ephedrine, other (specify)." (from ACC key data elements) Cocaine (Class: causes or risk factors) → Cocaineabuse (UMLS CUI: C0009171) Sentence: "Mechanical circulatory support in chronic heart failure evolved as a means of supporting patients awaiting transplantation, and this indication provided successful transition to heart transplantation and enhanced post-transplantation outcomes." (from Branwald) <i>Transplantation (Class: treatment)</i> → Heart transplantation (UMLS CUI: C0018823)
2. Separate terms into two or multiple terms when they contain "and" or "or" and the combined terms cannot be mapped to the target terminologies.	<i>Weight gain or loss</i> → weight gain (UMLS CUI: C0043094); weight loss (UMLS CUI: C0043096) <i>Mitral, aortic, tricuspid, and/or pulmonic valve surgical replacement</i> → Replacement of aortic valve (UMLS CUI: C0003506); Replacement of mitral valve (UMLS CUI: C0026268); Replacement of tricuspid valve (UMLS CUI: C0190119); Replacement of pulmonary valve (UMLS CUI: C0190129).
3. Restrict UMLS terminology sources to SNOMED CT, LOINC, RxNorm, and MeSH for concept mapping. For terms that can be mapped to multiple UMLS concepts, choose the concept that contains the target sources (example a). Terms that can be mapped to UMLS concepts, but not to one of the target sources should be left unmapped (example b).	For example, (a) <i>Abdominal fullness</i> was mapped to Abdominal bloating (UMLS CUI: C1291077) instead of Fullness abdominal (UMLS CUI: C0235318) because the C1291077 has source of SNOMED CT. (b) <i>Acute dyspnea</i> was left unmapped instead of mapping to acute dyspnea (UMLS CUI: C0743323) because C0743323 is not contained in the target sources.
4. Use semantic types to choose a proper mapping when terms can be mapped to multiple concepts.	<i>For diagnostic tests or results</i> , preferred semantic types are: Laboratory procedure, Laboratory or test result, e.g., <i>Blood urea nitrogen</i> → Blood urea nitrogen measurement (UMLS CUI: C0005845) <i>Atrial fibrillation</i> → ECG: atrial fibrillation (UMLS CUI: C0344434) For <i>treatment</i> , preferred semantic types are: pharmacologic substance, therapeutic or preventive procedure, e.g., <i>Digitalis</i> → Digitalis preparation (UMLS CUI: C0304520) <i>Yoga</i> → Yoga (UMLS CUI: C1883583)
5. Map term to the UMLS concept as close in meaning as possible.	<i>Chest radiograph</i> → Plain chest X-ray (UMLS CUI: C0039985) <i>Salt restriction</i> → Low sodium diet (UMLS CUI: C0012169)
6. Map terms with modifiers (e.g., daily, severe) by post-coordinating multiple UMLS concepts.	<i>History of Chagas disease</i> → Chagas disease (UMLS CUI: C0041234) + medical history (modifier) <i>Daily serum electrolytes</i> → Serum electrolytes measurement (UMLS CUI: C0587355) + daily (modifier)

**Table 4**

Contribution of each knowledge source to the four classes and the final heart-failure vocabulary.

Sources	Classes				Final vocabulary (N = 648)
	Causes or risk factors (N=435)	Signs or symptoms (N = 233)	Diagnostic tests or results (N = 590)	Treatment (N=477)	
Harrison's	71(16%)	79 (34%)	75 (13%)	77 (26%)	298 (18%)
UpToDate	110(25%)	45(19%)	120(20%)	116(24%)	380 (23%)
ACC key data element	126 (30%)	47 (20%)	115(19%)	100(23%)	386 (23%)
ESC guideline	81(17%)	54 (23%)	205 (35%)	154(32%)	484 (29%)
DynaMed	162 (37%)	55 (24%)	118(20%)	148(31%)	474 (29%)
ACC guideline	194(45%)	63 (27%)	146(25%)	178(37%)	565 (34%)
Braunwald	150 (34%)	147(63%)	298(51%)	239 (50%)	810(49%)

**Table 5**

Top 5 frequently occurring heart failure concepts in four classes.

Classes	UMLS CUI/code	Concept	Frequency
<i>Signs or symptoms</i>	C0013404	Dyspnea	136
	C0268000	Body fluid retention	55
	C0018810	Heart rate	48
	C0546817	Fluid overload	47
	C0015672	Fatigue	46
<i>Causes or risk factors</i>	C0020538	Hypertension	133
	C0027051	Myocardial infarction	133
	C0004238	Atrial fibrillation	132
	C0010054	Coronary arteriosclerosisDiabetes mellitus	123
	C0011849	Diabetes mellitus	92
<i>Diagnostic tests or results</i>	C0428772	Left ventricularejection fraction	219
	C0232174	Cardiac ejection fraction Brain	195
	C1095989	natriuretic peptide measurement	134
	C0022662	Kidney function tests	110
	C0013516	Echocardiography	82
<i>Treatment</i>	C0003015	Angiotensin-converting enzyme inhibitors	471
	C0001645	Adrenergic beta-antagonists	402
	C0012798	Diuretics	313
	C0521942	Angiotensin II receptor antagonist	250
	C1167956	Cardiac resynchronization therapy	215



**Table 6**

The agreement of the concepts among seven knowledge sources.

Agreement score	Braunwald	ACC guideline	DynaMed	ESC guideline	ACC key data element	UpToDate
Braunwald						
ACC guideline	0.45					
DynaMed	0.37	0.42				
ESC guideline	0.45	0.43	0.34			
ACC key data element	0.30	0.32	0.31	0.31		
UpToDate	0.43	0.46	0.38	0.40	0.30	
Harrison's	0.41	0.41	0.38	0.40	0.28	0.43

**Table 7**

Examples of heart-failure terms of four class types that occurred in one, two, three and seven sources.

Examples	Terms from 1 source	Terms from 2 sources	Terms from 3 sources	Terms from 7 sources
<i>Signs or symptoms</i>	Narrow pulse pressure; Jaw pain; Mottled skin; Presyncope; Purple-blue nail bed	Exercise intolerance; Abdomen distended; Abdominal pain; Apex beat displaced; Ventricular filling pressure increased	Reduced exercise tolerance; Early satiety; Respiratory distress; Dyspnea on exertion; Hypoxia	S3–third heart sound; Orthopnea; Fatigue; Dyspnea; Angina pectoris
<i>Causes or risk factors</i>	Deep vein thrombosis; Dysglycemia; Collagen diseases; Cilostazol; Egg consumption	Hyperlipidemia; Chronic kidney failure; Pulmonary arterial hypertension; Viral myocarditis	Fabry disease; Angina, unstable; Familial cardiomyopathy; Beriberi; Rheumatic fever	Coronary arteriosclerosis; Myocardial ischemia; Myocardial infarction; Diabetes mellitus; Hypertension
<i>Diagnostic tests or results</i>	Measurement of liver enzyme; Blood cell count; Hemoglobin Alc measurement; Indirect bilirubin measurement; Oral glucose tolerance test	Blood pressure monitoring; Myocardial biopsy; ECG: left ventricular strain; Pharmacologic and exercise stress test; Prolonged QRS duration	Cardiovascular monitoring; Cardiopulmonary exercise test; ECG: atrial fibrillation; calcium measurement; Blood pressure monitoring	Maximum oxygen uptake; Radionuclide ventriculography; Fluid overload; Magnetic resonance imaging; Left ventricular ejection fraction
<i>Treatments</i>	Alcohol deterrents; Sheng-Mai San; Dietary supplementation; Repair of pulmonary valve; Genetic counseling; Cell therapy	Ablation of atrioventricular node; Epoprostenol; Pericardiectomy; Weight reduction regimen; Hydralazine hydrochloride	Mitral valvuloplasty; Implantation of ventricular assist device; Fluid intake restriction; Implantation of CRT-D; Pneumococcal vaccination	Angiotensin II receptor antagonist; Inotropic agent; Heart transplantation; Diuretics; Angiotensin-converting enzyme inhibitors