Medical Terminologies in Action

A Report to the Board of Scientific Counselors
September 2017

Kin Wah Fung, MD, MSc, MA, FRCSEd

Computer Science Branch

Contents

1. Background

A medical terminology is a set of terms that standardize the recording of clinical findings, interventions, circumstances and events to support clinical care, decision support, research, quality improvement and other healthcare related activities. The basic function of medical terminologies is to enumerate all the terms that will be used in a certain domain (a controlled vocabulary). Many terminologies go beyond this to provide some form of organization (the commonest is a hierarchical structure), definitions and relationships between the terms. Medical terminologies are used for information capture, storage, exchange and retrieval in electronic health records (EHR). They allow efficient and unambiguous communication and sharing of medical information. They are integral to data interoperability and are a key enabler of an integrated nationwide health information system that promises increased patient safety and reduced cost. In the age of data-centric biomedical research, medical terminologies facilitate the integration of heterogeneous data sources to support big data analytics.

The National Library of Medicine (NLM) has a long history of supporting and conducting research, infrastructure development and policy studies to promote the design and deployment of medical terminologies. Since 1986, the Unified Medical Language System (UMLS) has been the flagship of NLM's efforts to promote the creation of more effective biomedical information systems through better and more innovative use of medical terminologies. On the policy level, NLM plays a leading role in U.S. government efforts to designate key health data standards as nationwide standards, to support the ongoing maintenance and free dissemination of important clinical terminology standards, and to promote and enable efforts to make health data standards more useful and usable in the U.S.

2. Project Objectives

This report covers multiple projects that fall into two broad categories. Firstly, research and development projects that advance the knowledge about medical terminologies and deliver resources that facilitate and promote the use of terminology standards. These projects include the CORE Problem List Subset of SNOMED CT, maps between SNOMED CT and ICD classifications, and RxTerms. Secondly, research projects that provide insights into how informatics can improve clinical care. Even though these projects are not primarily terminology projects, medical terminologies play a central role in the research methodologies. These projects include the Surescripts study, extraction of information from drug labels, and the comparison of commercial drug-drug interaction knowledge bases.

3. Project Significance

My research has led to the creation of terminology resources that are heavily used in the implementation of SNOMED CT for the problem list and RxNorm for medications in the EHR. The maps between SNOMED CT and ICD classifications are being implemented in clinical information systems and are useful for the integration of clinical and administrative data sets to support big data analytics. The Surescripts project shows that electronic prescription databases can significantly enhance manually collected medications data. The work on drug labels has successfully applied natural language processing to extract structured information from drug labels to assist the FDA's indexing initiative. The study on commercial drug-drug interaction knowledge bases documents their differences and points to possible ways of harmonization between knowledge sources.

4. CORE Problem List Subset of SNOMED CT

4.1 Motivation

SNOMED CT is the emergent international health terminology standard for encoding information such as clinical problems and procedures in the EHR. According to the Meaningful Use of EHR incentive program of the Centers for Medicare & Medicaid Services (CMS), the problem list of certified EHRs has to be encoded using SNOMED CT. SNOMED CT is a very comprehensive terminology. It has over 110,000 concepts that represent clinical problems. The typical size of problem list vocabularies in most institutions is below 30,000 terms. The goal of the CORE Problem List Subset is to identify the most commonly used SNOMED CT concepts to facilitate SNOMED CT implementation in problem lists. The CORE Subset originates from the UMLS-CORE study, which investigates the use of controlled vocabularies for **c**linical **o**bservations **r**ecording and **e**ncoding in specific sections of the EHR including the problem list, discharge diagnosis and reason for encounter. [1]

4.2 Methods and procedures

The UMLS-CORE project started in 2007 and collected problem list data sets from seven large scale health care institutions: Kaiser Permanente, Mayo Clinic, University of Nebraska Medical Center, Hong Kong Hospital Authority, Intermountain Healthcare, Regenstrief Institute and Beth Israel Deaconess Medical Center. The data set from the Veterans Health Administration was acquired in 2012. We analyzed the usage pattern of the problem list terms and found that the distribution of usage was highly skewed in all data sets, with about 20% of unique terms accounting for 95% of usage frequency on average. We focused on the most frequently used terms that accounted for 95% of usage in each institution, and mapped them to SNOMED CT through normalized lexical matching to the UMLS supplemented by manual review. The CORE Problem List Subset of SNOMED CT is the union of the mapped SNOMED CT concepts from all institutions.

4.3 Results and publications

The first CORE Subset was published in 2009 with 5,182 SNOMED CT concepts. The subset is updated quarterly to synchronize with new releases of SNOMED CT and the UMLS. The updates reflect changes in SNOMED CT e.g., newly added or retired concepts, new mapping of local terms and addition of new source data sets. The latest release of the CORE Subset contains 6,156 concepts, and is derived from data collected from eight large health care institutions covering a total of 18 million patients.

Findings of the usage pattern and overlap between the problem list data sets was published in JAMIA. [1] While there is modest overlap (around 20%) of problem list terms between institutions, the shared terms are used eight times more often than terms unique to an institution. We believe the CORE Subset will be useful as a 'starter set' to build a local problem list vocabulary. Not only will a common list save development time and effort, it will also reduce the variability and enhance the interoperability of problem list data.

We subsequently carried out an assessment of the potential coverage of the CORE Subset when applied to a new environment, and a projection of the growth of the subset if future data sets become available. We estimate that the CORE Subset will cover 80% of the frequently used terms, corresponding to 84% of the usage frequency, in a new institution. We project that the subset will reach a plateau of approximately

8,000 concepts with additional data sets, which is still relatively small and manageable compared to the whole of SNOMED CT. The study was published in JAMIA. [2]

We also did an experimental evaluation of the performance of the CORE Subset as a problem list vocabulary in a simulated data entry environment, comparing it to a clinical subset of SNOMED CT (100,000 concepts) and a problem list vocabulary used in the Mayo Clinic (24,358 terms). Despite its small size, the coverage of the CORE Subset is comparable to the other two problem list vocabularies. With the CORE Subset, the users can find their terms more quickly, which is probably due to the shortness of the pick list returned. [3]

4.4 Impact

The primary goal of the CORE Subset is to support implementation of SNOMED CT in the EHR. Based on the 2017 UMLS user annual reports, 600 UMLS licensees have used the CORE Subset, and among them 309 have implemented SNOMED CT in an EHR. The CORE Subset is not expected to be exhaustive and cover every concept that will ever be needed. The user has to augment their own problem list vocabulary by adding concepts outside the CORE Subset according to local requirements. However, the CORE Subset gives a significant head start by identifying many of the concepts that are known to be useful.

In addition to the EHR, the CORE Subset has found other use cases. The CORE Subset identifies a relatively small collection of clinical concepts whose importance is substantiated by actual usage data. Compared to the whole SNOMED CT, the CORE Subset is a more manageable target and stands as a proxy for the study of the clinical content in SNOMED CT. In 2010, SNOMED International (previously IHTSDO) did a comprehensive review of 100 concepts from the CORE Subset in an internal quality assessment exercise. The CORE Subset has also been used by other researchers as a tool for SNOMED CT quality assurance, [4-6] inter-terminology mapping [7] and terminology research [8-13] activities.

5. Mapping between SNOMED CT and ICD classifications

5.1 SNOMED CT to ICD-10-CM Map

5.1.1 Motivation

While SNOMED CT is a clinical terminology, the ICD codes (e.g. ICD-9-CM, ICD-10-CM and ICD-10-PCS) are considered administrative classifications, which are designed for specific purposes such as statistical reporting and financial accounting. They often lack the coverage and granularity needed to support direct patient care. SNOMED CT is more suitable as an EHR terminology than the ICD classifications, as reflected in the recommendations of the Meaningful Use program. However, due to the practical need of outputting ICD codes for reimbursement and other administrative functions, many EHRs use the ICD classifications as the primary encoding terminology. To encourage EHRs to use SNOMED CT to encode clinical information, we have created a map from SNOMED CT to ICD-10-CM. The map will enable EHRs that capture clinical information in SNOMED CT to generate ICD-10-CM codes in a quasi-automatic way and avoid the need for repeated coding. A map between SNOMED CT and ICD classifications will also allow integration of data from disparate sources such as EHRs and claims databases for data analytics research.

5.1.2 Creation of the SNOMED CT to ICD-10-CM map

5.1.2.1 Methods and procedures

To map from SNOMED CT to ICD-10-CM is to create a link between two very different artifacts: a clinical terminology and an administrative classification. It is not always possible to find a one-to-one correspondence between a SNOMED CT concept and an ICD code. In some cases, more than one ICD code is required to fully encode the meaning of a SNOMED CT concept. For example, the SNOMED CT concept *301011002 Escherichia coli urinary tract infection* requires two ICD codes used together: *N39.0 Urinary tract infection, site not specified* and *B96.20 Unspecified Escherichia coli as the cause of diseases classified elsewhere*. In addition, there are cases in which more than one possible ICD codes can be used for the same SNOMED CT concept, and the correct code depends on the patient's age, gender, co-morbidities and other factors. For example, *73430006 Sleep apnea* can map to either *P28.3 Primary sleep apnea of newborn* or *G47.30 Sleep apnea, unspecified* based on whether the patient is a neonate or adult. To allow the map to reflect the coding guidelines in ICD-10-CM, we use a rule-based approach to represent different map targets contingent on age, gender and co-morbidities.

We have recruited map specialists who possess knowledge about SNOMED CT and ICD classifications and trained them on the mapping tools and methodology. To ensure consistency and quality of the map, we employ the "dual independent mapping" method. Every mapping has to be corroborated by two independent sources. Most concepts are mapped independently by two map specialists and compared. Only if they agree will the mapping be finalized. Otherwise a third specialist will review and reconcile the difference. For a small number of concepts, there is an external map resource available (e.g. the donation from Kaiser Permanente). The mapping created by one map specialist is first compared to that source. If they agree the mapping is finalized. If they disagree, the external map will be disregarded and the concept is mapped by another map specialists, following the workflow described above.

5.1.2.2 Results and publications

All SNOMED CT concepts from the three hierarchies: Clinical finding, Event and Situation with explicit context are considered in scope for mapping. This includes a total of about 120,000 concepts. We started with concepts that would more likely be used in the EHR, such as the CORE Problem List Subset, and published the first map in 2012 containing 15,613 concepts. Over the years, we have regularly expanded the coverage the map and updated it for new releases of SNOMED CT and ICD-10-CM. The latest release in March 2017 covered 100,185 concepts. We are on course to cover the full scope of the map in the next release.

To showcase how the map can be used in an interactive way by a clinician, we developed a web-based tool called I-MAGIC (**I**nteractive **M**ap-**A**ssisted **G**eneration of **I**CD **C**odes; https://imagic.nlm.nih.gov). The user can search for a SNOMED CT term by typing in keywords, and at the click of a button ICD-10-CM codes will be suggested together with options to refinement based on laterality, episode of care etc.

This project has generated several publications covering our mapping methods, approach to handle age specification in ICD-10-CM and how the two maps, SNOMED CT to ICD-10-CM (created by NLM) and SNOMED CT to ICD-10 (created by SNOMED International) have learned and benefited from each other. [7, 14, 15]

5.1.3 Evaluation of the benefits of using the map

5.1.3.1 Methods and procedures

Since the primary goal of the SNOMED CT to ICD-10-CM map is to generate administrative codes, we carried out an experiment to evaluate the potential benefits of using the map to support coding. A collaborator from Nebraska University Medical Center supplied anonymized clinical notes which included chief complaints, medical history, physical examination, medications, diagnostic assessments and plan. In addition, all problem list terms picked by the clinician, together with the corresponding SNOMED CT encoding (generated by the EHR vendor) were included. Based on the SNOMED CT problems, we used our SNOMED CT to ICD-10-CM map to find candidate ICD-10-CM codes. Four professional coders applied codes to each note through a secured website, with or without seeing the map-suggested ICD-10-CM codes. We hypothesized that the availability of the map-suggested codes would make coding more efficient, consistent and accurate. The outcome measurements were coding time, inter-coder reliability and coding accuracy measured by recall, precision and F-score using a gold standard established by a Delphi consensus process.

5.1.3.2 Results and publications

For coding efficiency, there was significant reduction in average coding time from 9 to 7.5 minutes per clinical note. There was also modest improvement in coding reliability and accuracy. Less experienced coders generally benefitted more from the map suggestions than more experienced ones. Altogether, there were 553 codes in the gold standard and 390 codes (70.5%) were included in the candidate codes suggested by the map. We did a failure analysis for the cases in which the correct code was not found by our map. Most of these cases were attributable to issues not related to the map per se. In 39% of the cases, the relevant problem term was missing. In 21% of cases, a suboptimal problem term was entered which could not have led to the required ICD-10-CM code. For 28% of cases, the proprietary map from the problem term to SNOMED CT was suboptimal. Only in 12% of the cases was the failure caused the SNOMED CT to ICD-10-CM map itself. After adjusting for the factors unrelated to the map, the accuracy of the map-suggested codes is comparable to the average human coder. The findings will be submitted to an informatics journal.

5.1.4 Impact

The SNOMED CT to ICD-10-CM map is a heavily used terminology resource. Based on the 2017 UMLS user annual reports, 2,357 users are using the map. Among them are health care institutions, terminology knowledge base vendors and EHR vendors. For example, Partners HealthCare System reports in a publication their use of the map to facilitate the adoption of ICD-10-CM and to help clinicians retrieve and identify the most precise ICD-10-CM codes for a given clinical encounter. [16] Another example is the Indian Health Service (IHS) who uses the map in their EHR system to generate ICD-10-CM codes based on SNOMED CT concepts picked by clinicians. The rules in our map are used to refine the ICD-10-CM codes. We are currently collaborating with the IHS to explore the possibility of making their resources available to other users of the map.

The map and the I-MAGIC application was acknowledged in an open letter from the National Committee on Vital and Health Statistics to the Department of Health and Human Services in relation to the transition from ICD-9-CM to ICD-10-CM:

*"[HHS should] take this opportunity of converting the ICD classification system from ICD-9-CM to the ICD 10 code sets to align this rule with the Meaningful Use Rule that specifies SNOMED CT as the standard clinical terminology for coding diagnoses on the problem list...... [Clinicians] would then have the capability of mapping SNOMED CT codes to ICD-10-CM using **national standardized tools, such as the one recently developed by the National Library of Medicine (NLM) called I-Magic**. This will ensure that each coding standard is used for the purpose for which it was designed, thereby helping to mitigate the ICD-10-CM user interface challenges. **The NLM-developed I-Magic tool is a good example of a national, standardized user-friendly interface tool** for the conversion from clinical language to these structured terminologies and classifications (SNOMED CT, ICD-10-CM)."*

Other SNOMED International member countries who are using ICD-10-CM have also been using the SNOMED CT to ICD-10-CM map. Belgium has directly contributed to this project by providing a mapping specialist for one year to expedite the completion of the map.

5.2 ICD-9-CM to SNOMED CT map

5.2.1 Motivation

Many EHR systems still contain clinical information encoded in ICD-9-CM. To facilitate EHR systems to use SNOMED CT as the primary encoding terminology, it is desirable that the legacy ICD-9-CM data be translated to SNOMED CT. Through code translation it is possible to merge newly collected data with historic data, and the EHR can also benefit from the information in SNOMED CT for clinical decision support and other functions. The goal of the ICD-9-CM to SNOMED CT maps (two separate maps for diagnosis and procedure) is to facilitate the translation between the two coding systems.

5.2.2 Methods and procedures

We obtained lists of commonly used ICD-9-CM diagnosis codes (11,848 codes) and procedure codes (3,645 codes) from CMS. We created candidate maps based on the UMLS and the SNOMED CT to ICD-9-CM map created by SNOMED International. We did limited manual review for maps with a small number of SNOMED CT target codes and try to create one-to-one maps if possible, since the one-to-one maps are most useful for automatic code translation. However, not all ICD-9-CM codes could have one-to-one mapping to SNOMED CT. For example, there are the "catch-all" codes (i.e. the NEC or Not Elsewhere Classified codes, e.g., *480.8 Pneumonia due to other virus not elsewhere classified*) that could not be mapped to a single SNOMED CT concept. For such cases, it is not possible to translate an ICD-9-CM code to a SNOMED CT concept without human intervention.

5.2.3 Results and publications

We published the first version of the maps in 2012. We have been updating the maps yearly to synchronize with changes in SNOMED CT (ICD-9-CM has been frozen since 2013). In 2016, we obtained a map between ICD-9-CM and SNOMED CT by Nova Scotia Department of Health and Wellness which we used to validate and expand the scope of the maps. In the latest release, the diagnosis map provides maps to 11,221 ICD-9-CM codes of which 7,906 (70%) are one-to-one maps. The procedure map has maps for 2,241 ICD-9-CM procedure codes of which 1,753 (78%) are one-to-one maps.

5.2.4 Impact

Based on the 2017 UMLS user annual reports, 1,513 users are using either the diagnosis or the procedure map from ICD-9-CM to SNOMED CT.

5.3 SNOMED CT to ICD-10-PCS map

5.3.1 Motivation

ICD-10-PCS is a brand-new procedure coding system created by the U.S. Centers for Medicare and Medicaid Services (CMS) through a contract with 3M Health Information Systems. In 2015, ICD-10-PCS replaced the ICD-9-CM procedure codes for in-patient procedure coding in the U.S. Having a map between SNOMED CT and ICD-10-PCS will facilitate ICD-10-PCS coding, promote reuse of SNOMED CT coded clinical data and facilitate integration of clinical and administrative data to support data analytics.

5.3.2 Methods and procedures

I am leading a project group at SNOMED International to explore ways to create a map between SNOMED CT and ICD-10-PCS. This project group has representatives from Belgium and Spain. The group has studied various ways of automatically mapping between the two terminologies. These methods include lexical matching of the ICD-10-PCS index, indirect mapping through the General Equivalence Map (GEMs, published by CMS) and ontological alignment between the SNOMED CT attributes and ICD-10-PCS axes.

5.3.3 Results and publications

Each automatic mapping method has its strengths and limitations. For example, lexical mapping of the ICD-10-PCS index gives accurate maps but only applicable to a small number of concepts. Ontological mapping has broader coverage but the accuracy is lower. We have published on the findings of these studies. [17, 18]

5.3.4 Impact

This project is still in the exploratory phase. No mapping data has been published yet. Through publications and presentations, we have attracted attention from terminology service vendors and data analytics companies who have participated in our meetings. With more ICD-10-PCS data becoming available, the demand for a map to SNOMED CT will likely grow in the future.

6. RxTerms

6.1 Motivation

The development of RxTerms was prompted by a practical need for a good interface terminology to input medication information. The use cases then were the NLM's Personal Health Record project and CMS's post-acute care assessment tool. While RxNorm is the U.S. drugs reference terminology standard, the normalized drug names in RxNorm are long and not designed for use in data entry.

6.2 Methods and procedures

RxTerms separates the full names in RxNorm into two parts: drug name + route (e.g., INDERAL (Oral-pill)) and strength + dose form (e.g., 80 mg tab) to avoid big pick lists with long and hard-to-read names.

We exclude drugs that are obsolete or unavailable in the U.S. RxTerms also incorporates other user-friendly features such as synonyms and abbreviations (e.g., HCTZ for hydrochlorothiazide) and "tall man" lettering recommended by FDA to avoid medication errors (e.g., ChlorproMAZINE and ChlorproPAMIDE).

6.3 Results and publications

The first release of RxTerms was in 2008. We have been producing monthly updates to synchronize with each monthly release of RxNorm. The latest release includes 10,486 generic drugs, 7,879 branded drugs, 272 generic drug packs and 389 branded drug packs. New features and data elements introduced to RxNorm are incorporated in RxTerms when applicable.

We did an evaluation of RxTerms and found that it covers 99% of both generic and brand names of 200 most commonly prescribed drugs in the U.S. Compared to using RxNorm drug names, data entry using RxTerms is more efficient, requiring fewer keystrokes and returning shorter pick lists. [19]

6.4 Impact

On the RxTerms website, we request users who download the RxTerms data files to voluntarily submit their name, affiliation and email. To date, we have 2,073 registered users, but some of them may not be active users. Every year, the list grows by about 250 names. RxTerms can also be accessed through the RxNav APIs. In 2016, 25 million RxNav queries were directed to RxTerms. Based on user feedback and questions, RxTerms is used by hospital information systems, drug-related websites, software vendors and mobile apps developers.

7. The use of an electronic prescription database (Surescripts) for medication reconciliation in the Emergency Department

7.1 Motivation

Medication history is an essential part of patient assessment, particularly in emergency care. It provides clues about patient problems and information needed for removing, adjusting or adding therapies. On average, it takes 9 to 24 minutes to gather a patient's medication history. Despite the effort, most medication histories are incomplete. Patients do not know the name or the strength of many of their medications, and they forget some altogether. Direct inquires to pharmacies uncover medications not reported by the patient in up to 53% of hospitalized patients. [20] In a systematic review, there was omission in 10% to 61% of the patient medication histories across 22 studies. [21]

Today, prescription records for almost every prescription and refill reside in large umbrella computer systems – especially those of pharmacy benefits managers (PBM) and payers. In principle, hospitals and care providers could tap these central computer systems to get a complete history of prescriptions filled for all of their patients someday. One example of this kind of information resources is Surescripts. With over 40 collaborating PBMs, payers and pharmacies, Surescripts provides this pharmacy derived information to care providers as a real time service. They provided the information for free to clinicians in regions struck by Hurricane Katrina. At the time of this study the members of this collaborative were processing 2.5 billion US prescriptions per year, representing 60-70% of the US prescriptions covered by

commercial insurance. However, in practice, only a small number of hospitals and institutions are currently making use of such a valuable source of information.

The goal of this study is to raise awareness of the availability of information sources such as Surescripts and to provide some benchmarking of the value of such an electronic prescription database. Specifically, we want to know the proportion of patients that will be covered and the quality and completeness of the information delivered.

7.1 Methods and procedures

We did the study in the Emergency Department (ED) of Suburban Hospital, Bethesda MD, a 225 bed community hospital and regional trauma center with 45,000 visits per year. All patients attending the ED within the three months' study period were included. The study was reviewed and approved by the Office of Human Subjects Research (OHSR) at NIH and the Institutional Review Board of Suburban Hospital.

Suburban's ED nurses routinely collected a medication history from their patients and entered it into the hospital information system as part of their triage process. At the time of this study, medication names were entered as free text. To retrieve a patient's prescription records, Surescripts required the first and last name, birth date, zip code and gender. Surescripts then reported whether the patient was in its registry or not and delivered a record with the name, code, dispensing date, amount dispensed and prescriber's name for each prescription record it carried for that patient. All communication occurred automatically and electronically according to the Health Level Seven V.2 messaging standard.

To normalize the medications from both sources, we converted their drug names into RxNorm generic ingredient names by lexical matching supplemented with manual review. We used the FDA's compendium of generic OTC medication names to differentiate between over-the-counter (OTC) and prescription medications (PM). Furthermore, we used the National Drug File Reference Terminology (NDF-RT) to map the drugs to their classes (e.g. hypoglycemic, anticonvulsants).

We generated a list of active PM from the Surescripts report based on the fill date and amount prescribed, allowing a gap of 7 days before the ED visit to compensate for drug hoarding and skipped doses. We compared this list to the list of PM in the nurse's history (ED history). To estimate the clinical significance of the drugs revealed by the Surescripts records, we identified a set of high risk drugs with narrow therapeutic range e.g. hypoglycemic agents, anticonvulsants and anticoagulants. These drugs have been shown to account for a disproportionally large percentage of adverse drug events leading to ED visits and knowledge of them could potentially impact decisions in patient care.

7.3 Results and publications

In the study period, 9,426 unique patients visited the ED and were triaged by the triage nurse. According to the ED history 61% of patients reported at least one PM. Surescripts managed to match 65% of all patients, 53% of all patients had at least one PM at any time in the past year and 39% had at least one PM that was active. For the patients who had at least one active PM from Surescripts, the ED history was missing an average of 1.4 out of 5.4 active PMs (the union of the Surescripts active PM and ED history). The same quantity of PMs (but not the same drugs) was missing from the Surescripts active PM list. So the two sources complemented each other. The Surescripts active PM list delivered nearly as many of the patient's active medications as the more labor-intensive ED list, and enriched it by 28%, even after an 80% downward adjustment, assuming that not all prescribed drugs were actually taken by the patient.

Among the 14,929 PMs carried in the Surescripts active PM lists, 1,411 (9%) were high risk drugs, of which 229 PMs were not captured by the ED PM list. Overall, 985 unique patients (27%) had at least one high risk drug in their Surescripts active PM list, and in 209 unique patients (6%), at least one high risk drug missed by the ED PM list.

The electronic pharmacy prescription report has many advantages beyond its bolstering of the manual history. Firstly, it delivers unambiguous medication names compared to the manual ED history (for which we could not identify 3% of the names entered), and it always provides the dose form and strength, which were missing for 21% of drugs in the manual history. Secondly, it arrives in seconds rather than the minutes to hours (if one counts time to completion of pharmacist review), and busy ED personnel do not have to spend precious time to obtain it. Thirdly, its availability is not affected by patient consciousness or ED workload surges, and thus could be a godsend during disasters. Furthermore, the Surescripts active list contained high risk drugs among 11% of the patients who bypassed the nurse triage and 37% of those had some drug information from Surescripts - information of special importance to the critically ill. Finally, the full-year prescription record, presented as an easy-to-read flow sheet, could help identify medication compliance problems or drug seeking behavior (e.g. obtaining narcotic drugs from many providers).

The results are published in the Annals of Emergency Medicine. [22]

7.4 Impact

As far as we know, this is the first study on this topic. We hope that by raising the awareness of the availability and potential benefits of such information services, more health care providers can have access to electronic prescription data for routine patient care. Because of the study, we created an easy-to-read summary report of the Surescripts results and made it available to the ED clinicians at Suburban Hospital after the completion of the study.

8. Extraction of information from drug labels

8.1 Motivation

The effective use of clinical decision support (CDS) has been shown to improve the safety and quality of drug prescribing, resulting in significant reduction in medication errors. To provide timely advice during the drug prescribing process, the CDS function needs access to knowledge about drugs, such as therapeutic class, drug-drug interactions, adverse effects, indications and contra-indications. The availability of standardized drug knowledge bases has been identified as one of the critical elements that can help to realize the benefits of CDS. [23]

Drug knowledge bases are generally derived from three types of sources. Home-grown sources are laborious to create and maintain. Commercial sources can be expensive, and their use often implies "locking-in" to a specific company or vendor. Both home-grown and commercial knowledge bases often use proprietary rather than national or international data standards, and this hinders their interoperability and potential for sharing across institutions or practice settings. Publicly available knowledge bases do not have these drawbacks. However, the existing public resources are scarce and their contents limited.

Drug package inserts (drug labels) are a comprehensive, up-to-date and authoritative source of drug information which is publicly available. The drug labels for most prescription drugs and many over-the-counter drugs can be found on the DailyMed website collaboratively maintained by FDA and NLM. These electronic drug labels follow the SPL (structured product labeling) standard that enhances their machine readability. However, SPL only provides a structure to separate the drug label into sections (e.g. Indications & Usage, Contraindications, Precautions, and Adverse Reactions), the content within the sections is still in free narrative text. To unleash the knowledge in the drug labels, they need to be transformed into data encoded in standard terminologies, which can then be utilized by the inference engine in a CDS application. The goal of this research is to extract information from the drug labels using natural language processing and map it to standard terminologies.

8.2. Extraction of drug indications

8.2.1 Methods and procedures

We downloaded all drug labels for human drugs from DailyMed and extracted the Indications & Usage section based on the LOINC code in the SPL specification. We broke down the narrative text in the section into the smallest text segments identifiable by the XML tags (e.g. <section>, <paragraph>, <item>). Each segment was then submitted to the MetaMap API. The output from MetaMap was processed to remove negated concepts. Semantic type filtering was then done using the semantic group for disorders. We also used the list created by Fan and Friedman to refine the semantic type filtering for disorder concepts. [24] Finally, we excluded some high level UMLS concepts (e.g. disease, inflammation, infection) which were too general to be useful.

The drug labels could be linked back to RxNorm drug entities through their unique identifier (set-ID). We determined that the best level to represent drug indications was the Semantic Clinical Drug Form (SCDF) level (e.g. amoxicillin oral tablet). The ingredient level (e.g. amoxicillin) is potentially ambiguous as the indications for the same chemical compound could be different depending on the dose form (e.g. steroids as ointments, inhalants or tablets would have different indications). On the other hand, difference in strength would seldom affect indications, so we ignored strength information. To evaluate the accuracy of our method, the extracted indications for 300 commonly prescribed drugs were reviewed by seven physicians for accuracy. Each physician reviewed 60 labels, of which 20 were shared by all reviewers for assessment of inter-rater agreement.

We also compared our results to two external sources. Drug-indication pairs were extracted from the NDF-RT through the UMLS by the relationships "may_treat" and "may_prevent". The NDF-RT indications were specified mostly at the ingredient level and were propagated to the SCDF level through the RxNorm relationships. From the Semantic Medline project, we identified associations between drugs and the diseases they might treat. These associations were mostly at the ingredient level and were propagated to the SCDF level.

8.2.2 Results and publications

Altogether 6,797 labels (2,104 unique drugs) were processed. MetaMap extracted 19,473 indications of which 3,468 indications belonged to the 300 commonly prescribed drugs that were manually reviewed. Overall, the MetaMap-extracted indications had recall, precision and F-score of 0.95, 0.77 and 0.85 respectively. These was moderate agreement among reviewers, with pairwise kappa between 0.515 to 0.914, and overall kappa of 0.713.

Failure analysis revealed that word sense ambiguity (e.g. *bacterial* strain vs. *muscle* strain) was the commonest cause of error. The next common cause was extraneous information. For example, hypertension was picked up as the indication for potassium salt because of the sentence: "The use of potassium salts in patients receiving diuretics for uncomplicated essential hypertension is often unnecessary". Missed negations and high level concepts accounted for the remaining errors.

NDF-RT provided 42,507 indications for 6,554 drugs. Of these, 421 drug-indication pairs overlapped with those that were manually reviewed, and 409 were found to be correct (precision 97%). Semantic Medline yielded 696,297 indications for 6,074 drugs. Of these, 977 indications were manually reviewed and 905 found to be correct (precision 93%).

The study was published in JAMIA. [25]

8.3 Extraction of drug-drug interactions

8.3.1 Methods and procedures

For extraction of drug-drug interactions from drug labels, we used a supervised machine learning framework built on a support vector machine (SVM) and a shallow linguistic features space. The overall workflow is shown in this diagram.
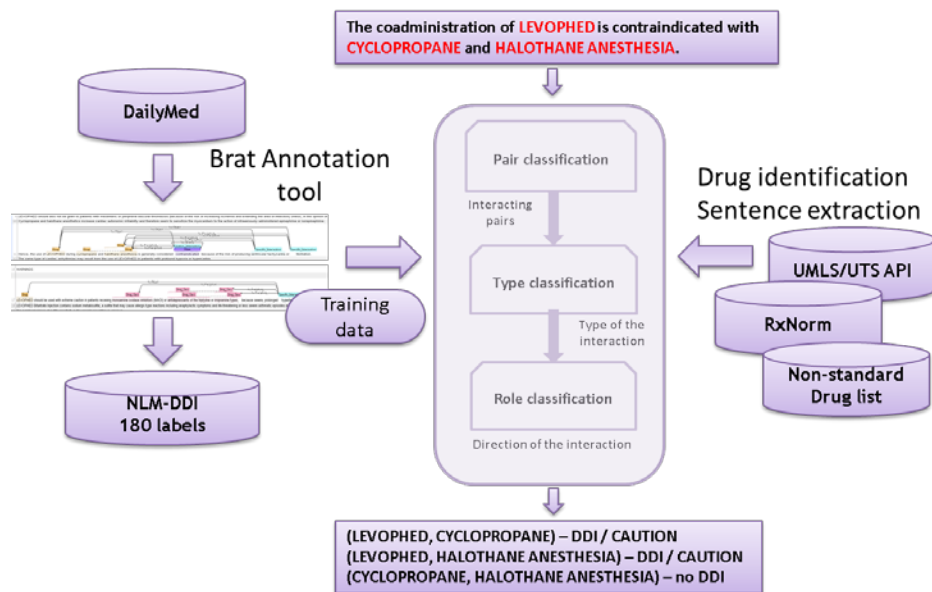
Development of the annotated corpus

Our FDA collaborators identified 180 cardiovascular drugs which we annotated according to our schema to identify pharmacologic entities (including drugs, drug classes and other substances such as grapefruit juice or alcohol), four types of interaction (general caution, increase or decrease of activity of the object drug, and interaction with specific effects) and role of the pharmacologic entities (object or precipitant drug). Annotation was done independently by two informatics specialists and the difference was reconciled by a third reviewer.

Building the machine learning pipeline

We used the MetaMap API to identify pharmacologic entities in sentences from eight sections of the SPL where DDI information may be present. Every sentence with two or more pharmacologic entities was retrieved for further processing. If more than two entities were present, each pairwise combination of the entities was considered individually. The classification process was divided into three sub-tasks: pair classification to determine whether there was a mention of DDI, type classification to determine the type of interaction, and role classification to determine the role of each entity.

Features used for the classifier included both global and local features. Global features divided the sentence into three sub-spaces according to the relation to the two pharmacologic entities (e.g. before, between or after the entities), and extracted the stems of words, n-grams of the stems of words, part-of-speech (POS) tags of words and n-grams of POS tags of words from each sub-space. Local features were features within a certain window size from each of the entities and included the original token, its stem, POS tag and orthographic class. We used LIBSVM with a linear kernel to perform the classification tasks.

Evaluation of the classifier

We divided the annotated data set into a 70% training and 30% test set. We performed a 10-fold cross validation on the training set to compute the cost parameter of the linear kernel. We evaluated the performance of the classifier for the three classification sub-tasks by recall, precision and F-score.

We used the SemEval Drug interaction Task 9 (2013) corpus in order to test the robustness of our feature space when applied to text other than drug labels. The SemEval corpus was composed of texts from DrugBank and MEDLINE, which we pooled together and used to retrain our classifier. We also used a reference list of high-priority DDIs (the ONC list) [26] to evaluate our classifier. The ONC list was a list of DDIs of high clinical significance created by a stakeholder panel commissioned by the Office of National Coordinator of Health IT (ONC). We identified drug labels related to drugs on the ONC list and applied our classifier to extract DDIs. The purpose of this test was to test our classifier for drug labels outside the cardiovascular drugs on which it was trained.

8.3.2 Results and publications

The annotation of the 180 drug labels was completed successfully with inter-rater agreement (kappa statistic) varying from 0.72 to 0.9 for different types of information annotated. The annotated corpus was made available through our website (http://lhce-brat.nlm.nih.gov/NLMDDICorpus.htm).

For pair classification, our classifier had recall, precision and F-score of 0.869, 0.818 and 0.842 respectively using combined global and local features. For type classification, the recall was 0.625 – 0.926, precision 0.833 – 0.918 and F-score 0.714 – 0.909 for different types of interaction. The specific interactions had the best performance while the decrease interactions had the worst performance, probably related to their paucity in the data set. For role classification, our classifier was able to achieve 100% accuracy.

Applying our classifier to the SemEval corpus, we were able to achieve recall 0.436 – 0.832, precision 0.703 – 0.891 and F-score 0.586 – 0.822 for the four sub-tasks. Our classifier outperforms most of the methods in the SemEval competition.

In the attempt to extract the ONC high-priority DDIs, our classifier was only able to retrieve 59 out of 360 DDIs in the ONC list. However, this poor performance could be explained because the majority of the missing DDIs were expressed at the drug class level, without specific mention of the drugs involved. If we propagated DDIs of the drug classes to their members using standard drug classification references (e.g. NDF-RT), we would be able to identify many of the missing DDIs in the ONC list.

This research is still ongoing. With the feedback from FDA domain experts, we will look for ways to improve the performance of our classifier.

8.4 Impact

In the indication extraction study we have shown that it is possible to use a publicly available natural language processing tool (MetaMap) to extract drug indications from drug labels with reasonable recall and precision. When other sources of information (like NDF-RT) are available to corroborate the results, precision can be improved significantly. Similar methods with very little modification can be used to extract adverse effects and contra-indications.

The JAMIA paper on the extraction of drug indications caught the attention of FDA who approached us for a collaborative research project to extract drug-drug interactions from drug labels using natural language processing. This led to the DDI extraction study described above. After a year of collaboration, FDA and NLM signed an interagency agreement through which FDA provides some funds to NLM to create an NLP pipeline to support FDA's SPL indexing initiative. The goal of the indexing initiative is to "transform [the content of drug labels] into discrete, coded, computer-readable data in order to advance patient safety, promote quality of health care delivery, and improve health care outcomes. Transforming the narrative text to structured information encoded in national standard terminologies is a prerequisite to the effective deployment of drug safety information in drug labels for clinical decision support". The output of the indexing initiative will eventually be made available to the public through the DailyMed website. The interagency agreement (renewable for five years) is currently in its second year.

9. Comparison of commercial drug-drug interaction knowledge bases

9.1 Motivation

DDIs represent a significant cause of adverse drug events. One study reported that DDI accounted for 17% of adverse drug reactions leading to hospitalization. [27] In another study, 4.4% of elderly patients received prescriptions with a risk of severe interactions. [28] DDIs are potentially preventable because of their predictability. A comprehensive, accurate and evidence-based knowledge base (KB) is a prerequisite to the effective deployment of a CDS system. Since the resources and expertise needed to develop and maintain a home-grown DDI KB are only available to a few large academic centers, most organizations choose to purchase their KBs from commercial vendors. However, studies have shown significant variability between DDI knowledge sources. Inconsistent evaluation and classification of interactions have been cited as factors contributing to excessive DDI alerts. The goal of this study is to systematically compare commercial DDI KBs known to be deployed in clinical environments.

9.2 Methods and procedures

We contacted five commercial DDI KB vendors and First Databank (FDB), Micromedex and Multum agreed to participate in our study. All three KBs ranked the DDIs according to the level of severity: contraindicated, major (or severe), moderate and minor. We excluded the minor DDIs (e.g. interactions with herbal remedies) because they were less important and tended to be less consistently represented. We mapped the drugs to RxNorm at the SCD (semantic clinical drug e.g., Azithromycin 500 MG Oral Tablet) and GPCK (generic drug pack e.g., 6 Pack of Azithromycin 500 MG Oral Tablet) levels which were closest to the forms used in the KBs. Strength information was often included because some DDIs were strength-dependent. Two KBs came with their own RxNorm mappings, and we mapped the remaining one to RxNorm through lexical matching supplemented by manual review.

A DDI could be represented in three ways: first, as a pair of clinical drugs, specifying the active ingredients, strength and dose form (e.g., trimipramine 100 mg capsule and albuterol 2 mg tablet); second, as a pair of ingredients (e.g., trimipramine and albuterol) and third, as a pair of drug classes (e.g., tricyclic antidepressants and sympathomimetics). We rolled up the DDIs in the KBs to the ingredient and drug class level and compared them at all three levels. Additionally, we tested the KBs against a high-priority list of DDIs (the ONC list). [26] Most of the DDIs in the ONC list were listed at both the drug class and ingredient levels. For two drug classes (tricyclic antidepressants and QT-prolonging drugs) without the corresponding ingredients, we expanded them to the ingredient level using standard references. We assessed the coverage of the ONC list by the three KBs. Finally, we used a prescription filling data set to assess the quantity of DDI alerts that would be generated by the KBs. The drugs in the prescription filling data set were mapped to RxNorm at the SCD and GPCK levels to match with the DDIs in the KBs.

9.3 Results and publications

The size of the KBs varied considerably in terms of clinical drug pairs. FDB had the least drug pairs (1.6 million), followed by Micromedex (4.5 million) and Multum (4.8 million). In all KBs, contraindicated DDIs were the smallest category and moderate DDIs the largest. Overall, the number of drug pairs that were commonly configured to generate interruptive alerts (contraindicated and major/severe categories) was 490,260 (30.8%), 2,311,324 (51.9%) and 468,822 (9.8%) for FDB, Micromedex and Multum respectively. Altogether, the three KBs contained 8.6 million unique drug pairs, of which 6.8 million (79.4%) were unique to one KB, 1.3 million (15.5%) were found in two KBs and 0.4 million (5%) in all three KBs. The percentage of unique unshared drug pairs (i.e., not found in any other KB) was 35.6%, 65% and 70.9% for FDB, Micromedex and Multum respectively.

Regarding the severity ranking of the DDIs, we found that there was less inconsistency among the KBs than suggested by earlier studies. On average, if the same clinical drug pair occurred in two KBs, their severity ranking agreed about 60% of the time. The severity rankings of contraindicated DDIs were more in agreement compared to the lower severities. Generally, the pattern of overlap at the ingredient pair level was similar to the clinical drug level. At the drug class level, both the degree of overlap and the agreement in severity ranking were considerably higher. Overall, 48.5% of DDIs at the class level were shared by all three KBs, much higher than the ingredient (8.7%) and clinical drug (5%) levels.

The 15 DDI rules in the ONC list expanded to 1,027 pairs of ingredients. Overall, FDB, Micromedex and Multum did not cover 106, 122 and 105 of the ONC pairs respectively. We asked the experts providing the KBs to review all of the missing pairs. Versioning (i.e., missing DDI was available in a newer version

of the KB) and obsolete drugs (drugs no longer in the market) accounted for most of the omissions. Overall, if we adjusted for the unintentional differences (versioning, obsolete drugs, mapping problem etc.), the coverage of the ONC list would become 98.7%, 98.8% and 99.9% for FDB, Micromedex and Multum respectively.

Our prescription filling dataset covered 1.9 million patients with 14 million prescriptions. Considering all severity levels, the alerts generated by FDB, Micromedex and Multum would be 163, 329 and 751 alerts per 1,000 prescriptions respectively. Counting only contraindicated and major/severe DDIs (which usually result in interruptive alerts), 25, 145 and 84 alerts per 1,000 prescriptions would be generated by FDB, Micromedex and Multum respectively. Applying the ONC list to the prescription dataset would generate 43,047 alerts (3 alerts per 1,000 prescriptions). The overwhelming majority (97.6%) of the ONC alerts was related to two DDI rules: co-prescription of statins and protease inhibitors, and two QT prolonging drugs. Overall, FDB, Micromedex and Multum covered 97.9%, 85.9% and 99.8% of the ONC alerts respectively. Adjusting for the unintentional differences (versioning, mapping problem etc.), the overall coverage of ONC alerts would become 99.8%, 99.9% and 99.9% for FDB, Micromedex and Multum respectively.

The findings were published in JAMIA. [29]

9.4 Impact
Through this study we found that there were considerable differences between three commercial DDI KBs. However, the differences were less pronounced for the more severe DDIs. There was very high coverage by all three KBs of the ONC DDIs and the alerts generated by the ONC list. A lot of the discrepancies could be traced back to differences in the definition of drug classes and their membership (e.g. QT prolonging drugs, CYP-450 inhibitors). A single authoritative source of these drug classes would help to reduce variability in DDI knowledge sources.

In view of the variability among different sources of DDI information, it has been suggested that an expert panel with a centralized organizer or convener should be established to develop and maintain a standard set of DDIs for CDS in the U.S., as has been done in the Netherlands. However, the intensive logistics and trend towards DDI customization at individual institutions may make this effort difficult to implement. The Pharmacy Quality Alliance (PQA) is convening Stakeholder Advisory Panels for the purpose of creating and maintaining a consensus-based minimum DDI data set. The findings of this study will hopefully shed some light on ways to harmonize between DDI knowledge sources.

10. Summary and future plan
Medical terminologies are at the heart of every modern EHR system. Without clinical information encoding through medical terminologies, many advanced functions such as clinical decision support, automatic reporting and intelligent data retrieval or aggregation will be difficult to implement. Data interoperability between disparate systems will not be possible without the shared semantics provided by medical terminologies. Data-driven research (e.g., the learning health care system proposed by the Office of the National Coordinator of Health IT) often requires the pooling of data from disparate sources which relies on mapping between different terminologies. In many medical informatics research dealing with raw data from clinical information systems or biomedical research, the normalization of data through

medical terminologies is an essential step. In short, medical terminologies can be considered as one of the foundation stones on which the discipline of medical informatics is built.

This report provides an overview of my research which revolves around medical terminologies. Through my research I have developed resources to promote and facilitate the adoption of terminology standards, and help users to reap more benefits from encoded clinical information (e.g., re-use of coded data through inter-terminology mappings). My clinical background and experience in building clinical information systems often lead me to ask questions about how informatics can improve patient care. The research that I embark on to answer these questions often requires in-depth knowledge about medical terminologies, be they clinical terminologies (e.g., SNOMED CT), drug terminologies (e.g., RxNorm) or administrative terminologies (e.g., ICD-9-CM, ICD-10-CM). Going forward, I shall continue to develop, maintain and improve the terminology resources, and to apply my expertise in medical terminologies and informatics skills to advance the knowledge on important clinical and informatics topics.

11. References

1.   Fung, K.W., C. McDonald, and S. Srinivasan, The UMLS-CORE project: a study of the problem list terminologies used in large healthcare institutions. J Am Med Inform Assoc, 2010. 17(6): p. 675-80.

2.   Fung, K.W. and J. Xu, An exploration of the properties of the CORE problem list subset and how it facilitates the implementation of SNOMED CT. J Am Med Inform Assoc, 2015. 22(3): p. 649-58.

3.   Fung, K.W., et al., Testing Three Problem List Terminologies in a simulated data entry environment. AMIA Annu Symp Proc, 2011. 2011: p. 445-54.

4.   Rector, A. and L. Iannone, Lexically suggest, logically define: quality assurance of the use of qualifiers and expected results of post-coordination in SNOMED CT. J Biomed Inform, 2012. 45(2): p. 199-209.

5.   Rector, A.L., S. Brandt, and T. Schneider, Getting the foot out of the pelvis: modeling problems affecting use of SNOMED CT hierarchies in practical applications. J Am Med Inform Assoc, 2011. 18(4): p. 432-40.

6.   Agrawal, A., et al., The readiness of SNOMED problem list concepts for meaningful use of electronic health records. Artif Intell Med, 2013. 58(2): p. 73-80.

7.   Campbell, J.R., et al., Semantic Interoperation and Electronic Health Records: Context Sensitive Mapping from SNOMED CT to ICD-10. Stud Health Technol Inform, 2013. 192: p. 603-7.

8.   Nadkarni, P.M. and J.A. Darer, Migrating existing clinical content from ICD-9 to SNOMED. J Am Med Inform Assoc, 2010. 17(5): p. 602-7.

9.   Steindel, S.J., A comparison between a SNOMED CT problem list and the ICD-10-CM/PCS HIPAA code sets. Perspect Health Inf Manag, 2012. 9: p. 1b.

10.   Lopez-Garcia, P., et al., Usability-driven pruning of large ontologies: the case of SNOMED CT. J Am Med Inform Assoc, 2012. 19(e1): p. e102-9.

11.   Abdoune, H., et al., Assisting the translation of the CORE subset of SNOMED CT into French. Stud Health Technol Inform, 2011. 169: p. 819-23.

12.   Lamy, J.B., et al., A Semi-automatic Semantic Method for Mapping SNOMED CT Concepts to VCM Icons. Stud Health Technol Inform, 2013. 192: p. 42-6.

13.   Hogan, W.R. and V.N. Slee, Measuring the Information Gain of Diagnosis vs. Diagnosis Category Coding. AMIA Annu Symp Proc, 2010. 2010: p. 306-10.

14.   Xu, J. and K.W. Fung, Handling age specification in the SNOMED CT to ICD-10-CM cross-map. AMIA Annu Symp Proc, 2012. 2012: p. 1014-22.

15.   Fung, K.W. and J. Xu, Synergism between the mapping projects from SNOMED CT to ICD-10 and ICD-10-CM. AMIA Annu Symp Proc, 2012. 2012: p. 218-27.

16.   Cartagena, F.P., et al., Leveraging the NLM map from SNOMED CT to ICD-10-CM to facilitate adoption of ICD-10-CM. J Am Med Inform Assoc, 2015. 22(3): p. 659-70.

17.   Fung, K., et al., Leveraging lexical matching and ontological alignment to map SNOMED CT surgical procedures to ICD-10-PCS. AMIA Annu Symp Proc 2016, 2016: p. 570-579.

18.   Fung, K.W., et al., Achieving Logical Equivalence between SNOMED CT and ICD-10-PCS Surgical Procedures. AMIA Annu Symp Proc, 2017. (accepted for presentation as paper).

19.   Fung, K.W., C. McDonald, and B.E. Bray, RxTerms - a drug interface terminology derived from RxNorm. AMIA Annu Symp Proc, 2008: p. 227-31.

20.   Mersfelder, T.L. and R.J. Bickel, Inpatient medication history verification by pharmacy students. Am J Health Syst Pharm, 2008. 65(23): p. 2273-5.

21.   Tam, V.C., et al., Frequency, type and clinical importance of medication history errors at admission to hospital: a systematic review. CMAJ, 2005. 173(5): p. 510-5.

22.   Fung, K.W., et al., Comparison of electronic pharmacy prescription records with manually collected medication histories in an emergency department. Ann Emerg Med, 2013. 62(3): p. 205-11.

23.   Teich, J.M., et al., Clinical decision support in electronic prescribing: recommendations and an action plan: report of the joint clinical decision support workgroup. J Am Med Inform Assoc, 2005. 12(4): p. 365-76.

24.   Fan, J.W. and C. Friedman, Semantic classification of biomedical concepts using distributional similarity. J Am Med Inform Assoc, 2007. 14(4): p. 467-77.

25.   Fung, K.W., C.S. Jao, and D. Demner-Fushman, Extracting drug indication information from structured product labels using natural language processing. J Am Med Inform Assoc, 2013. 20(3): p. 482-8.

26.   Phansalkar, S., et al., High-priority drug-drug interactions for use in electronic health records. J Am Med Inform Assoc, 2012. 19(5): p. 735-43.

27.   Pirmohamed, M., et al., Adverse drug reactions as cause of admission to hospital: prospective analysis of 18 820 patients. BMJ, 2004. 329(7456): p. 15-9.

28.   Rosholm, J.U., et al., Polypharmacy and the risk of drug-drug interactions among Danish elderly. A prescription database study. Dan Med Bull, 1998. 45(2): p. 210-3.

29.   Fung, K.W., et al., Comparison of three commercial knowledge bases for detection of drug-drug interactions in clinical decision support. J Am Med Inform Assoc, 2017.

13. Acknowledgements

Comparison of commercial knowledge bases
Dr Olivier Bodenreider
Joan Kapusnik-Uner (First Databank)
Jean Cunningham (Micromedex)
Stefanie Higby-Baker (Multum)

Administration of terminology research databases
Joseph Chow