**THE LISTER HILL NATIONAL CENTER
FOR BIOMEDICAL COMMUNICATIONS**

A research division of the U.S. National Library of Medicine

A Report to the Board of Scientific Counselors
April 2009

Applied Medical Terminology Research

Kin Wah Fung, M.D.

**U.S. National Library of Medicine, LHNCBC
8600 Rockville Pike, Building 38A
Bethesda, MD 20894**

Table of Contents

1. Background 3

2. Project Objectives 3

3. Project Significance 4

4. RxTerms – an interface terminology to RxNorm 4

4.1 Clinical drug names in RxNorm 4

4.2 The RxTerms solution for efficient data entry 6

4.2.1 Smaller lists and shorter names 6

4.2.2 Pruning of drugs 7

4.2.3 Adding user-friendly names 7

4.3 Evaluation of RxTerms 8

4.3.1 Coverage of generic and brand names 8

4.3.2 Data entry efficiency 9

4.4 Future work 11

5. Problem List Vocabularies 11

5.1 Methods 12

5.2 Results 13

5.2.1 Characteristics of the datasets 13

5.2.2 Usage pattern 14

5.2.3 Mapping to the UMLS 17

5.2.4 Overlap between the datasets 20

5.3 The CORE subset 21

5.3.1 Candidates of the CORE subset 21

5.3.2 Comparison of the candidate subsets 23

5.4 Future work 26

6. Inter-terminology Mapping 26

6.1 The need for inter-terminology mapping 26

6.2 Mapping algorithms explored 27

6.2.1 Semantic mapping (IntraMap) 27

6.2.2 Lexical mapping (MetaMap) 27

6.2.3 Combined semantic and lexical mapping 27
6.2.4 Evaluation of the algorithms 27
6.3 Results 28
6.3.1 Semantic mapping alone 28
6.3.2 Lexical mapping alone 28
6.3.3 Overlap between semantic and lexical mapping 29
6.3.4 Combined semantic and lexical mapping 29
6.4 Inter-terminology mapping in practice 31
7. The RxHub Project – a sneak preview 32
7.1 Background 32
7.2 Objectives and methods 32
7.3 Current status 33
8. Conclusion 33
9. Acknowledgements 34
10. References 35
11. Questions for the Board 38
Curriculum Vitae 39

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 enlist 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 facilitate information capture, storage, exchange and retrieval in electronic health records. They facilitate 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.

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 and services, through better and more innovative use of medical terminologies. The purpose of the UMLS is to improve the ability of computer systems to understand and manipulate biomedical meaning and to use this understanding to retrieve and integrate relevant machine-readable information for users. In its latest release (2008AB), the UMLS incorporates 126 biomedical terminologies and organizes 9 million names in these terminologies into 1.8 million concepts.

On the policy level, NLM has played 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 three research projects on medical terminologies: RxTerms, Problem List Vocabularies and Inter-terminology Mapping. These projects focus on practical issues, problems and barriers in the use of medical terminologies in computer systems. The research is often triggered by a specific need or challenge that arises when medical terminologies are deployed in real-life applications. The goal of this research is to facilitate and promote the use of standard medical terminologies, improve clinical documentation, enable efficient data reuse, enhance data interoperability and ultimately to improve patient care.

3. Project Significance

Medical terminologies are at the heart of every electronic health record system. Without information encoding enabled by medical terminologies, electronic medical records are little more than 'electronic page turners' which can only regurgitate data in the original form as entered by users. More advanced functions such as clinical decision support, automatic reporting and intelligent data retrieval or aggregation will not be possible. With the heightened interest of the new administration in the use of information technology in healthcare, and the pressing goal of a universal electronic health record by 2014, the importance of medical terminologies cannot be over-emphasized. By removing or lowering the barriers of acquisition and deployment of medical terminologies in clinical information systems, this research will contribute to the creation of a truly integrated and interoperable nationwide health information network.

4. RxTerms – an interface terminology to RxNorm

The development of RxTerms was triggered by a practical need for an interface terminology to capture medication information. In late 2007, the Lister Hill Center started development of NLM's Personal Health Record (PHR). There was a need for an efficient way for PHR users to enter their medications. Around the same time, NLM also assisted CMS (Centers for Medicare and Medicaid Services) in the development of their new assessment tool in the post-acute care environment (CARE). They also needed a drug interface terminology. Both applications were going to capture medication information encoded in a standard drug terminology. RxNorm was a natural pick for the standard, as it is the designated national standard for clinical drugs. The problem was to find an efficient way to capture information from the user and to transform that information into RxNorm codes.

4.1 Clinical drug names in RxNorm

RxNorm is a standardized nomenclature for clinical drugs created by NLM. A clinical drug is a pharmaceutical product given to (or taken by) a patient with a therapeutic or diagnostic intent. RxNorm serves as the bridge between different naming conventions used in disparate drug information systems. By creating a standardized set of drug names and linking them to equivalent names from various sources, RxNorm allows systems using different drug nomenclatures to share data efficiently at the appropriate level of abstraction.

There are two classes of concepts in RxNorm that represent clinical drugs: Semantic Clinical Drug (SCD) and Semantic Branded Drug (SBD). Every clinical drug is assigned an RxNorm unique identifier (RXCUI). The name of a clinical drug in RxNorm contains information about the ingredient(s), strength, intended route and dose form. In the case of a branded drug, the brand name is appended within square brackets. For example:

198440 Acetaminophen 500 MG Oral Tablet (SCD)
209459 Acetaminophen 500 MG Oral Tablet [Tylenol] (SBD)

Within RxNorm, generic and branded drugs are linked to each other and to the names of their individual components by a well-defined set of named relationships.

The initial attempt to build the data entry interface was to use the RxNorm clinical drug names directly as the interface terminology. When the user typed in a drug name, a pick list with all matching RxNorm clinical drug names would be displayed. This approach turned out to be unsatisfactory. Firstly, the pick lists were often big which made them difficult to display on the screen. Secondly, the RxNorm names were long and each name contained a lot of information, resulting in cognitive overload. It was difficult for users to select the right one. For example, when the user typed in 'amoxicillin', the pick list consisted of 49 names, most of them over 30 characters in length:

Amoxicillin 60 MG/ML Oral Suspension
Amoxicillin 400 MG/ML Injectable Solution
Amoxicillin 167 MG/ML Injectable Solution
Amoxicillin 200 MG/ML Injectable Solution
Amoxicillin 50 MG/ML Oral Suspension
Amoxicillin 100 MG/ML Oral Suspension
Amoxicillin 125 MG Chewable Tablet
Amoxicillin 200 MG Chewable Tablet
Amoxicillin 250 MG Oral Capsule
Amoxicillin 250 MG Oral Tablet
Amoxicillin 400 MG Chewable Tablet
Amoxicillin 80 MG/ML Oral Suspension
Amoxicillin 500 MG Oral Capsule
Amoxicillin 500 MG Oral Tablet
Amoxicillin 875 MG Oral Tablet
Amoxicillin 25 MG/ML Oral Suspension
Amoxicillin 10 MG/ML Oral Suspension
Amoxicillin 40 MG/ML Oral Suspension
Amoxicillin 125 MG/ML Oral Suspension
Amoxicillin 200 MG Oral Tablet
Amoxicillin 400 MG Oral Tablet
Amoxicillin 20 MG/ML Oral Suspension
Amoxicillin trihydrate 600 MG Disintegrating Tablet
Amoxicillin 250 MG / Clavulanate 125 MG Oral Tablet
Amoxicillin 125 MG / Clavulanate 31.2 MG Chewable Tablet
Amoxicillin 875 MG / Clavulanate 125 MG Oral Tablet
Amoxicillin 250 MG Chewable Tablet
Amoxicillin 5 MG/ML / Clavulanate 2.5 MG/ML Oral Solution
Amoxicillin 50 MG/ML / Clavulanate 10 MG/ML Injectable Solution
Amoxicillin 500 MG / Clavulanate 125 MG Oral Tablet
Amoxicillin 25 MG/ML / Clavulanate 6.25 MG/ML Oral Suspension
Amoxicillin 250 MG / Clavulanate 62.5 MG Chewable Tablet
Amoxicillin 200 MG / Clavulanate 28.5 MG Chewable Tablet
Amoxicillin 400 MG / Clavulanate 57 MG Chewable Tablet
Amoxicillin 50 MG/ML / Clavulanate 12.5 MG/ML Oral Suspension
Amoxicillin 40 MG/ML / Clavulanate 5.7 MG/ML Oral Suspension
Amoxicillin 80 MG/ML / Clavulanate 11.4 MG/ML Oral Suspension
Amoxicillin 25 MG/ML / Clavulanate 25 MG/ML Oral Suspension
Amoxicillin 50 MG/ML / Clavulanate 25 MG/ML Oral Suspension
Amoxicillin 400 MG / Clavulanate 28.5 MG Chewable Tablet
Amoxicillin 1000 MG / Clavulanate 62.5 MG Extended Release Tablet
Amoxicillin 120 MG/ML / Clavulanate 8.58 MG/ML Oral Suspension
12 HR Amoxicillin 1000 MG / Clavulanate 62.5 MG Extended Release Tablet
Amoxicillin 800 MG / Clavulanate 125 MG Oral Tablet
Amoxicillin 200 MG/ML / Clavulanate 28 MG/ML Oral Suspension

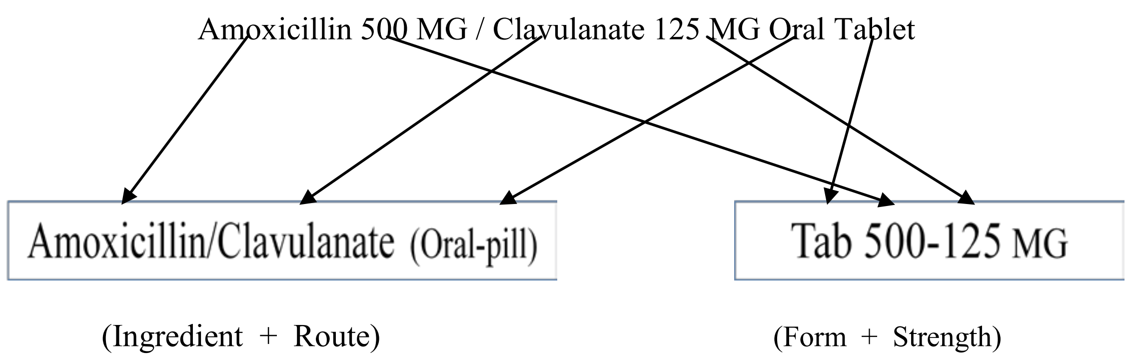
Amoxicillin 100 MG Oral Tablet
Amoxicillin 50 MG Oral Tablet
Amoxicillin 775 MG Extended Release Tablet
Amoxicillin 120 MG/ML / clavulanate potassium 8.58 MG/ML Oral Suspension

4.2 The RxTerms solution for efficient data entry

4.2.1 Smaller lists and shorter names

After some experimentation, it was found that both problems (the size of the pick lists and the length of the drug names) could be solved by segmenting and reorganizing the information contained in a clinical drug name.

The RxNorm name is broken down into two parts, the first part consists of [ingredient(s) + route] and the second part [form + strength].



The data entry is done in two steps. When the user types in ‘amoxicillin’, the pick list shows the matching ingredient and route combinations, which has only 6 choices:

Amoxicillin (Injectable)
Amoxicillin (Oral-liquid)
Amoxicillin (Oral-pill)
Amoxicillin/Clavulanate (Oral-liquid)
Amoxicillin/Clavulanate (Oral-pill)
Amoxicillin/Clavulanate XR (Oral-pill)

After the user picks one of them (Amoxicillin/Clavulanate (Oral-pill)), the available form and strength combinations will be shown:

Chewable Tabs	125-31.2 MG
Chewable Tabs	200-28.5 MG
Chewable Tabs	250-62.5 MG
Chewable Tabs	400-57 MG
Tabs	250-125 MG
Tabs	500-125 MG
Tabs	875-125 MG

The smaller lists and shorter names are much easier for users to pick from.

4.2.2 Pruning of drugs

RxNorm is a reference terminology which needs to be comprehensive and archival. On the other hand, the purpose of RxTerms is to facilitate efficient capture of medication information e.g. electronic prescription writing. There are drugs in RxNorm that are not likely to be useful in RxTerms and are therefore excluded. These include:

- Obsolete drugs – drugs that are flagged as obsolete in RxNorm.
- Non-US drugs – drugs that are available in the U.S. should normally have NDC codes. All drugs not associated with NDC codes in RxNorm are presumed to be unavailable in the U.S. and are pruned. However, since RxNorm may not contain all NDC codes (even though it is already the largest collection publicly available), this pruning may not be 100% accurate.
- Generic drug names with three or more ingredients – users are unlikely to type in the name of every ingredient when they refer to these drugs. Most commonly the brand name will be used instead. For example, a user is more likely to type in 'Cortisporin' than 'Bacitracin/Hydrocortisone/Neomycin/Polymyxin'. This step is particularly important for the pain and cold medications. Without the pruning, typing in drugs like 'acetaminophen' or 'chlorpheniramine' will return hundreds of drugs.
- Brand names containing the words 'aspirin' or 'acetaminophen' - these are mostly over-the-counter pain and cold medications that are not useful in prescribing anyway. Some names may actually be confusing e.g. Aspirin-free is acetaminophen and not aspirin.
- Allergenic extracts – not likely to be used in prescriptions
- Brand names containing the phrase 'brand of' e.g. 'Geneva brand of amiodarone hydrochloride'

4.2.3 Adding user-friendly names

4.2.3.1 Synonyms and abbreviations

There are synonyms and abbreviations that prescribers prefer to use such as ASA (aspirin), APAP (acetaminophen), INH (isoniazid) and HCTZ (hydrochlorothiazide). These names are not available in RxNorm. They are acquired from external sources and added to RxTerms. One source of drug synonyms is the National Ambulatory Medical Care Survey data file published by the NCHS (National Center for Health Statistics). The file contains drug names that are close to the way in which they are entered by prescribers, including some synonyms or abbreviations. Some drug synonyms and abbreviations are gleaned from the internal vocabulary table used by the Regenstrief Institute. Some synonyms are suggested by individual physicians involved in the project.

4.2.3.2 'Tall Man' lettering

The FDA has requested manufacturers of sixteen look-alike name pairs to voluntarily revise the appearance of their established names in order to minimize medication errors resulting from look-alike confusion. A convention called 'Tall Man' lettering is recommended which highlights the difference between drug names that are similar e.g. ChlorproMAZINE and ChlorproPAMIDE, CycloSPORINE and CycloSERINE. RxTerms has adopted this convention in the names of these drugs.

4.2.3.3 Insulin names

Insulins are among the most commonly prescribed drugs. There are 41 different generic forms of insulin in RxNorm. Insulins that are derived from animal sources are no longer available in the U.S. and are therefore suppressed. For the other insulins, the long and heterogeneous RxNorm names are replaced by 16 shorter and more prescriber-friendly names:

Insulin analog, Aspart (Injectable)
Insulin analog, Aspart Mixed 70/30 (Injectable)
Insulin analog, Detemir (Injectable)
Insulin analog, Glargine (Injectable)
Insulin analog, Glulisine (Injectable)
Insulin analog, Lispro (Injectable)
Insulin analog, Lispro Mixed 50/50 (Injectable)
Insulin analog, Lispro Mixed 75/25 (Injectable)
Insulin, human Lente (Injectable)
Insulin, human Mixed 50/50 (Injectable)
Insulin, human Mixed 70/30 (Injectable)
Insulin, human NPH (Injectable)
Insulin, human Regular (Injectable)
Insulin, human Regular U500 (Injectable)
Insulin, human Ultralente (Injectable)
Insulin, human, rDNA origin (Inhalant)

4.2.3.4 Concentration of liquid drugs

In RxNorm, all liquid dose concentrations are normalized to /ML when the standardized name is created. (e.g. Amoxicillin 50 MG/ML Oral Suspension). In clinical practice, most oral liquid drugs are prescribed in multiples of 5 ML (occasionally 15 ML). It will feel more natural to prescribers if liquid dose concentrations are expressed in this way. In RxTerms, oral liquid medication concentrations are shown as multiples of 5 (or 15) ML (e.g. Amoxicillin 250 MG/5ML).

4.3 Evaluation of RxTerms

4.3.1 Coverage of generic and brand names

To evaluate the coverage of RxTerms we used a list of 200 most common prescriptions in the U.S. called the RxList. The list contained both branded and generic names.

Of the 165 generic drug names on the list, all but one could be found in RxTerms. Among the ones that were found, 33 generic names did not have exact matches in RxTerms either because of minor naming variations (e.g. Folate vs. Folic acid) or different ordering of ingredients (e.g. promethazine / codeine vs. codeine / promethazine). The only missing generic drug was Divalproex. Divalproex sodium is the USAN (United States Adopted Name) equivalent of the INN (International Nonproprietary Name) name Valproate semisodium, and Valproate was present in RxTerms.

Among the 222 branded drug names, 206 were found in RxTerms. Among them, 7 were not exact lexical matches because of minor naming differences (e.g. TRI-LEVLEN vs. TRI

LEVLEN). Among the 16 that were not found in RxTerms, 13 were oral contraceptive pill packs. At the time of the evaluation, RxNorm did not contain pill packs. Since then, new drug classes have been added so RxNorm and RxTerms now cover names of oral contraceptives and other pill packs. Of the 3 missing brand names that were not oral contraceptives, Duragesic and Esidrix were not found in RxTerms but Duragesis and Esidrex were found. This might be a problem of common variations in spelling. The brand name Actonel was initially not found but it was subsequently added in later RxNorm releases.

The overall coverage of RxTerms was 99% (164/165) for generic drug names and 99% (206/209) for branded drug names that were not oral contraceptives.

4.3.2 Data entry efficiency

To evaluate the data entry efficiency of RxTerms, we simulated a prescription writing environment in which the user typed in the drug name one letter at a time. The input string was matched against the drug display names (ingredient + route) in RxTerms. We assumed that 'auto-completion', a popular feature in many form-filling applications, was used. This meant that stepwise matching occurred as the user typed in each additional letter. For example, when the user typed in 'dil' the returned list contained 'dilantin, dilaudid, dilocaine, dilor …' but as she typed in 'dila' the list shortened to 'dilantin, dilaudid, dilacor…'; and on typing 'dilan' only 'dilantin' remained. We only used the generic and brand names from the most prescribed drugs list that had exact matches in RxTerms because our evaluation algorithm could not identify non-exact matches. For each drug, we recorded the size of the lists as each additional letter was typed. We also noted the minimal number of keystrokes needed to return a list of less than 15 items. This was somewhat arbitrary but 15 items could usually be displayed without scroll-bars. For comparison, we repeated the process by matching against all non-obsolete generic and brand drug names in RxNorm.

Figure 1 shows the relationship of the median pick list size for all drugs to the number of letters typed in (keystroke). List sizes were generally big when less than 4 letters were typed (e.g. median list size for generic drugs at 3 keystrokes: RxNorm 361, RxTerms 82) so they were not included in the graph. The median list size for generic drugs was considerably smaller at every keystroke for RxTerms (yellow line) compared to RxNorm (blue line). The biggest difference was seen when 4 and 5 letters were typed (4 letters: RxNorm 48, RxTerms 9; 5 letters: RxNorm 28, RxTerms 4). For branded drugs, the difference was still seen but less pronounced.

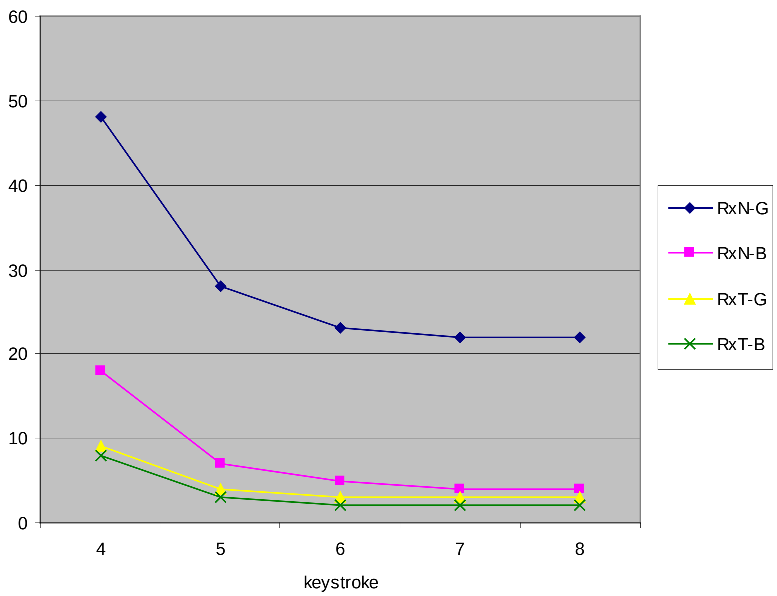


Figure 1. Median size of the returned lists in relation to the number of keystrokes (RxT: RxTerms, RxN: RxNorm, G: generic drugs B: branded drugs)

KS	RxT-G	RxN-G	RxT-B	RxN-B
≤ 3	15 (11%)	2 (2%)	38 (19%)	22 (11%)
≤ 4	86 (66%)	20 (15%)	133 (67%)	89 (45%)
≤ 5	115 (88%)	36 (27%)	184 (92%)	160 (80%)
≥ 6	128 (98%)	44 (34%)	199 (100%)	194 (97%)
Never	3 (2%)	87 (66%)	0 (0%)	5 (3%)
Total	131 (100%)	131 (100%)	199 (100%)	199 (100%)

Table 1. Efficiency evaluation for commonly prescribed drugs (KS: keystroke, RxT: RxTerms, RxN: RxNorm, G: generic drugs B: branded drugs)

Table 1 shows the minimal number of keystrokes needed to return a list of less than 15 items. There were some drugs for which even when the full name was entered the returned list was still 15 or above ('never'). Overall, using RxTerms for generic drugs, one could get a list of less than 15 choices within 5 keystrokes for 88% of drugs. This was considerably better than using RxNorm names (27%). The percentage of generic drugs that never got below 15 items was much lower for RxTerms (2%) compared to RxNorm (66%). The three drugs that failed to get below 15 were niacin, hydrochlorothiazide and potassium chloride. For branded drugs, the difference between RxNorm and RxTerms was smaller. Using RxTerms, one could get to below 15 in 5 keystrokes for 92% of branded drugs (vs. 80% for RxNorm). There were no branded drugs that failed to get below 15 for RxTerms (vs. 3% for RxNorm).

4.4 Future work

RxTerms is currently being used in the demonstration project of the CARE tool of CMS. It will also be used in the NLM's PHR. RxTerms has been released on the NLM website for public testing and feedback since November 2008 (http://wwwcf.nlm.nih.gov/umlslicense/rxtermApp/rxTerm.cfm). Users need to agree to a simple user agreement but no licensing is required. Monthly updates are provided to synchronize with the monthly full releases of RxNorm. Up to the date of this report, there are about 100 registered users of RxTerms. A survey of the users is planned to find out how they are using RxTerms, and to obtain user feedback. This will provide guidance to the future development of RxTerms.

5. Problem List Vocabularies

The problem list is a powerful way to organize and communicate clinical data and reasoning. It provides a convenient summary of the patient's active problems and significant co-morbidities. This information helps to facilitate the continuity of care, formulation of plan of treatment or further investigations and management of risk factors. The problem list has been recommended as an essential feature of electronic patient record systems. The way in which problem lists are generated varies between institutions. There are generally three ways to populate a problem list: as free text only, limited to some controlled vocabulary or a combination of both. To fully reap the benefits of an electronic problem list (e.g. patient-specific decision support, automatic generation of billing codes), the use of a controlled vocabulary is necessary. This research is focused on the use of controlled vocabularies in problem lists.

In the majority of electronic medical record systems, clinical narration (e.g. discharge summary, transfer notes) is still entered as free text. The problem list is usually the first and only part of the clinical content that is encoded by controlled vocabulary. Despite some efforts to develop a standardized problem list vocabulary, so far no standards have emerged. Most institutions create their own local vocabulary. Due to the need to generate ICD9CM codes for billing or public health reporting, many of these local vocabularies are originally derived from ICD9CM descriptions. However, being designed for statistical reporting of mortality and morbidity, the ICD9CM terms often do not extend well to cover clinical narratives which physicians wish to use. Compared to ICD9CM, SNOMED CT is more comprehensive and clinically oriented. Under the International Health Terminology Standards Development Organization (IHTSDO), SNOMED CT is poised to become the international standard for clinical documentation. There are some institutions that use SNOMED CT as the basis of their problem list vocabulary. Problem list vocabularies are seldom static. There are always requests for new terms to satisfy specific user needs. These new local terms which are not found in standard terminologies are called local extensions. As more local terms are added, the problem list vocabularies diverge more from the reference terminology and each other.

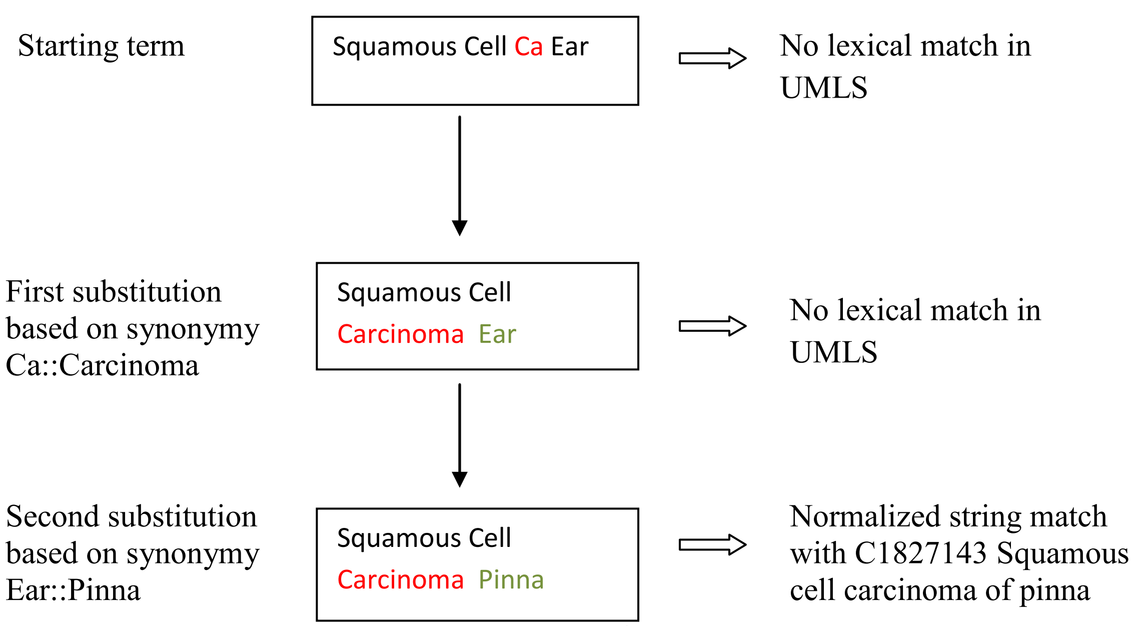
The problem list vocabularies research has two goals:

1. To study the problem list vocabularies of large health care institutions and to characterize them in terms of their size, pattern of usage and the extent to which they overlap with each other. This will provide insight into the potential barriers of information sharing across institutions
2. To identify a CORE (Clinical Observations Recording and Encoding) subset of the UMLS that can act as a basis (or 'core') from which problem list vocabularies are developed. This will help to reduce the variability among problem list vocabularies and enhance data interoperability

5.1 Methods

Health care institutions were contacted to see whether they used controlled vocabularies for their problem list or not. If they did, they were asked to share their lists for this research. In addition to the list of terms, the actual frequency of usage of each term was also collected. As far as possible, patient-based (instead of encounter-based) usage data was used to ensure uniformity for comparison and to remove the bias in favor of chronic problems that required repeated encounters. If available, any map from the local terminology to standard terminology was also collected to facilitate the mapping of the local terms to the UMLS.

Before the local vocabularies could be compared, the local terms were first mapped to the UMLS. Only exact (or synonymous) matches between the local terms and UMLS concepts were used. The process of mapping to the UMLS was done sequentially. The first step was lexical matching. The local terms were compared to English UMLS strings to look for exact, case-insensitive and normalized string (generated by the NORM program that comes with the UMLS lexical tools) matches. A further extension of lexical matching made use of a list of synonymous words. Local terms that did not have exact, case-insensitive or normalized matches were parsed to see if they contained words on the synonyms list. If so, the word was substituted with its synonym and lexical matching was repeated. If there were still no matches, another round of synonym substitution and matching was done. An example of lexical matching after synonym substitution is shown below.



The second step of the mapping made use of the local maps to standard terminologies if they were present. Some institutions labeled their maps explicitly to indicate whether a particular map was an exact match or not. Such exact maps were trusted as correct and the local term was mapped to the UMLS through their map target in the standard terminology (either ICD9CM or SNOMED CT). In cases where maps were not labeled explicitly as exact or not, the maps were manually inspected and only exact maps were used for mapping to the UMLS.

The final step was manual mapping. All terms that remained unmapped after the first two steps were manually mapped to the UMLS, using the UMLS RRF browser (distributed together with the UMLS) as the searching tool.

The resulting lists of mapped UMLS concepts were used for overlap analysis and creation of the CORE subset.

5.2 Results

5.2.1 Characteristics of the datasets

Datasets from six health care institutions were collected. They included Kaiser Permanente (KP), Mayo Clinic (MA), Intermountain Health Care (IH), Regenstrief Institute (RI), University of Nebraska Medical Center (NU) and Hong Kong Hospital Authority (HA). HA was the only non-U.S. institution. The patient population served by HA is primarily hospital inpatients. On discharge, the physician chooses one or more discharge diagnosis from the list of diagnostic terms, which is the list studied here. The function of the discharge diagnosis list is very similar to the problem list: a synopsis of the patient's problems, reminder for subsequent caregivers and to generate ICD codes for public health reporting. The characteristics of these datasets are summarized in Table 2.

	HA	IH	KP	MA	NU	RI
Patient population	inpatient	mixed	mixed	mixed	mixed	mixed
Patient count (million)	1.3	0.36	10	1.5	0.5	0.16
Period of data retrieval	3 years	snapshot	snapshot	3 years	snapshot	1 year
Total diagnosis count (million)	4.1	1.1	52	10	2.7	0.66
Average entry per patient	3.1	3.0	5.2	6.8	5.3	4.2
Total unique terms	12,449	5,685	26,890	14,921	13,126	3,166

Table 2. Characteristics of the datasets

Apart from HA, all the other institutions are mixed in- and out-patient facilities. The total number of patients covered by the datasets was 13.8 million. The data retrieval period ranged from 1 to 3 years. For IH, KP and NU, the datasets represented a snapshot of all patients in their systems. Except for MA, all datasets were patient-based data, meaning that if the same problem was recorded more than once for the same patient, it was counted only once. For MA, the data was encounter-based. However, since the Mayo Clinic is a tertiary referral center, the proportion of patients with chronic recurrent problems is small and this should not constitute a significant bias. The average number of problems per patient ranged from 3 to 7. Another noteworthy observation is that the size of the local vocabularies vary considerably, ranging from 3,166 to 26,890 terms, a difference of more than 8 fold.

5.2.2 Usage pattern

For each dataset, the pattern of usage was analyzed. We wanted to know how evenly (or unevenly) usage was spread across all terms. For a certain percent of usage, we calculated the percent of terms needed to cover that percent of usage (Figure 2). We used the percent of terms instead of the absolute number because the vocabularies differed considerably in size.

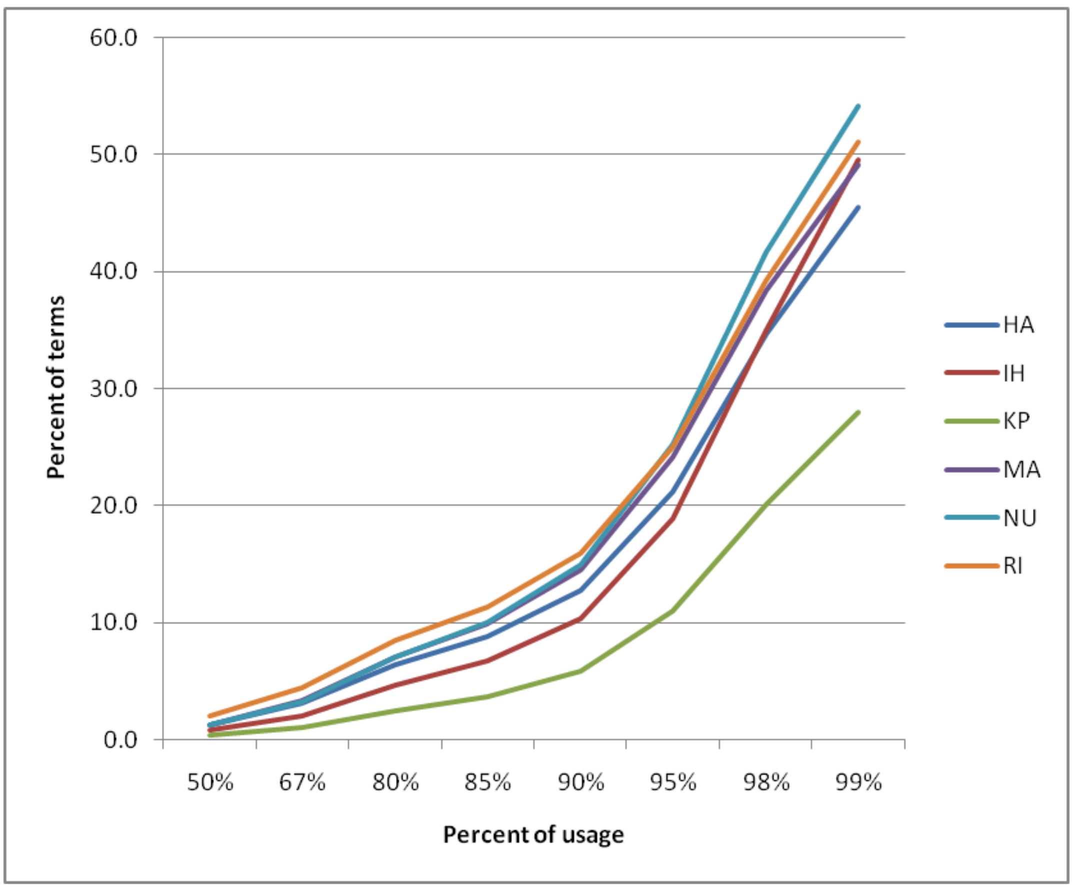


Figure 2. Usage pattern graph

There is a considerable degree of similarity of the usage patterns for 5 of the 6 institutions. KP seems to be the 'outlier' which can be explanation. The KP list was still undergoing significant expansion at the time the data was pulled. The recently added terms were used much less frequently because they were available only for a short period of time. This explains why the percentage of terms being used in KP was considerably lower than the other datasets.

The question still remains as to why there is similarity in the usage patterns. Since local vocabularies are created independently, used in different settings and by different users, one would not expect them to have similar usage patterns. One possible explanation is that the datasets are somehow connected by a common characteristic that they share - their inherent relationship to ICD9CM codes. For the 4 U.S. institutions (IH, MA, NU and RI), ICD9CM codes are needed for billing. For HA, ICD9CM codes were historically needed for public health reporting (now ICD10 codes are used). For this reason, it is likely that these local vocabularies were initially closely linked to the ICD9CM classification. Even with the subsequent evolution of these vocabularies, it is conceivable that the overall usage patterns still somehow echo the epidemiological pattern of diseases as classified by ICD9CM codes. To find further support for this hypothesis, we looked at the National Hospital Discharge Survey (NHDS) data published by the National Center for Health Statistics (NCHS). The 2004 NHDS collected data from 270,000 inpatient records in 500 U.S. hospitals. Among the data collected was the discharge diagnosis coded in ICD9CM. In the 2004 NHDS data, a total of 8,659 unique ICD9CM codes were used to code 1.7 million diagnosis (average of 6.5 discharge diagnosis per patient). When the usage pattern of the ICD9CM codes in NHDS was superimposed on the usage patterns of the datasets, the NDHS usage pattern lies almost in the center of the cluster of the 5 institutions (Figure 3).

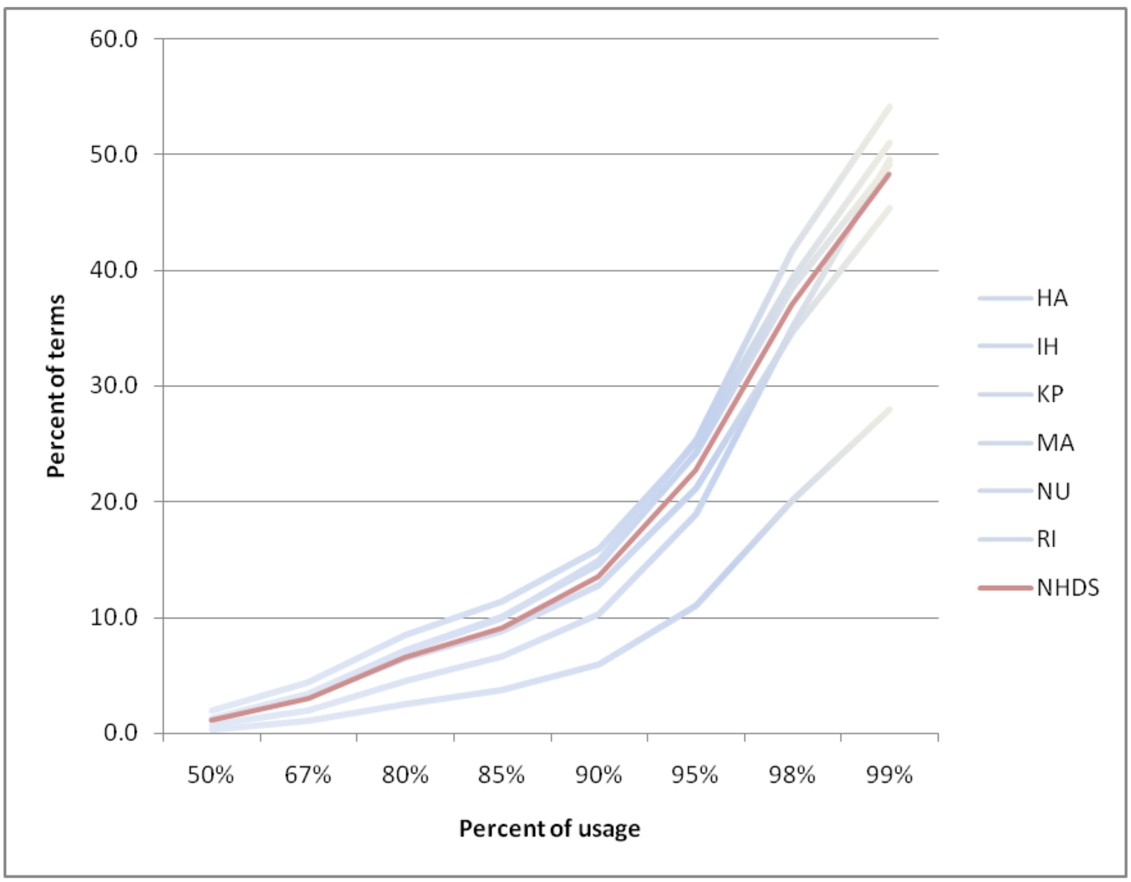


Figure 3. Usage pattern of ICD9CM codes from NHDS data (superimposed on the usage pattern of the local vocabularies)

Another possible explanation of the similarity of usage pattern is the Zipf's law. Zipf's law is an empirical law that predicts that many types of data studied in the physical and social sciences follow power law probability distributions. Zipf first observed this phenomenon in the usage of words in a corpus of natural language utterances. The frequency of any word is inversely proportional to its rank in the frequency table. Whether a certain dataset shows Zipfian distribution can be tested by plotting the data on a log-log graph, with the axes being log(rank order) and log(frequency). If the distribution is Zipfian the plot will be close to a straight line. Figure 4 shows the log-log plots for the datasets.

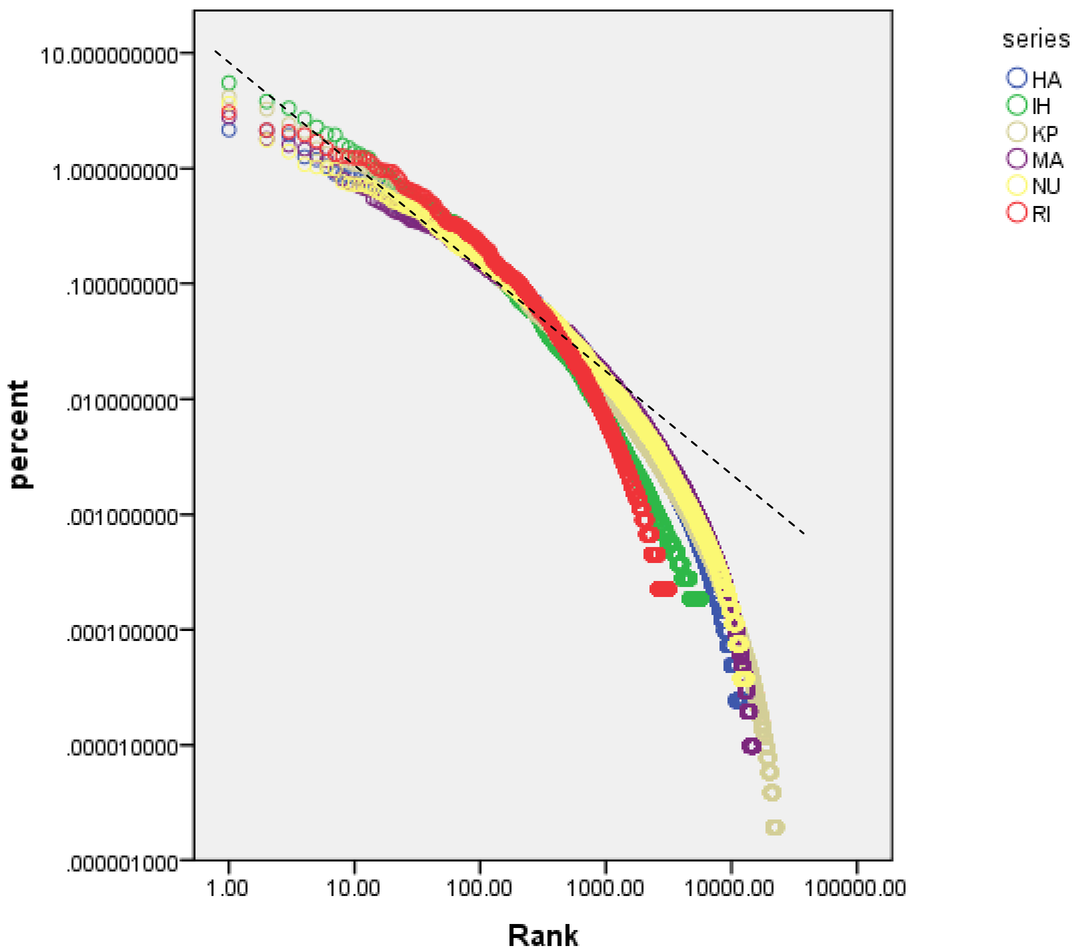


Figure 4. Log-log plot of usage percent vs. rank for all terms in the datasets

All the plots for the datasets show some degree of similarity. The initial portions of the plots are close to a straight line but the latter portions begin to deviate from it. However, this does not completely exclude a Zipfian distribution as it is not uncommon to observe this pattern even in datasets that are supposed to follow Zipf's law (e.g. the word frequency in Wikipedia). Plotted in this way, the KP dataset no longer deviates from the rest, as the newly added and rarely used terms are all concentrated in the lower portion of the line and do not affect its overall shape.

One further observation from the usage pattern is that one only needs a small proportion of the terms to get a high coverage of usage. The skew in distribution is even than that described by the Pareto principle (also known as 80-20 rule). If one concentrates on 20% of the terms, one already gets close to 95% of usage coverage.

5.2.3 Mapping to the UMLS

Altogether there were 76,237 terms from the 6 datasets. It would take a long time and considerable effort to map every term to the UMLS. Moreover, many of these terms were used very infrequently. Therefore, it was decided to map only the most frequently used terms covering up to 95% of usage. This reduced the number of terms to map to a more manageable size of 14,395 terms.

5.2.3.1 Results of mapping

Lexical mapping generated maps for 10,812 (75%) terms. The highest yield of lexical mapping came from exact matches (46%), followed by normalized string (14%) and case insensitive matches (10%). Synonymous word substitution added another 5% of maps.

Local maps resulted in maps for 1,007 (7%) terms, of which 700 (5%) were trusted maps and 307 (2%) were mapped after manual validation.

A total of 2,576 (18%) terms remained for manual mapping, which found maps for 1,442 (10%) terms. Overall, 1,134 terms (8%) could not be mapped to the UMLS.

	Number of terms	**Percent**
Lexical mapping	10,812	75%
Local maps	1,007	7%
Manual mapping	1,442	10%
Failed to map	1,134	8%
Total	14,395	100%

Table 3. Results of mapping to the UMLS

5.2.3.2 Unmappable terms

All terms that could not be mapped were categorized according to the reason for which they could not be found in the UMLS. The distribution of the various categories is shown in Table 4.

The commonest category of terms (more than half of the total) was highly specific and pre-coordinated terms such as:

Preterm labor after 22 weeks, before 37 weeks, w/o delivery, antepartum
DM 2 w diabetic end stage renal disease on dialysis
Slipped percutaneous transhepatic biliary drain catheter

Antepartum haemorrhage due to placenta praevia, type III

Reason for no UMLS map	Number of terms	% of no map terms
Highly specific	606	53%
Very general	120	11%
Administrative	84	7%
Laterality	81	7%
Negative finding	34	3%
Composite concept	32	3%
Meaning unclear	26	2%
Miscellaneous	81	7%
Total	1,134	100%

Table 4. Categories of terms that failed to map to the UMLS

Some terms were not found in the UMLS because they were more general than most diagnostic descriptions in standard terminologies, e.g.

Gyn Abnormality Other
Abnormal blood finding
Hx of major organ surgery
Hx of health hazards

Some terms were for administrative rather than clinical purposes, e.g.

No show
Doctors/nurses
Other Mr # exists
Administrative Visit
Advance directive not in chart
Administrative encounter for chart being opened in error

Some terms included laterality information. Most of these terms could be mapped without the laterality information, e.g.

Pain Heel Left
Renal stone, right
Bilateral renal stones

Terms from standard terminologies generally refer to positive findings. Terms that represented negative or unknown findings were not likely to be found in standard terminologies, e.g.

No ureteric stone
No urethral stricture

No abnormality detected in bladder
Rubella status unknown

Some terms combined two concepts which existed separately in the UMLS, but not together as a pre-coordinated concept, e.g.

Diarrhea with dehydration
Cognitive deficits from skull Fx
Vaccination for hepatitis B & hemophilus

Some terms could not be mapped because their meanings were unclear, e.g.

Conjunctiva Red
Infertility – error
DM 2, uncontrolled, w manifestation
Hx of stent placement

5.2.3.5 Relationship of usage to mappability

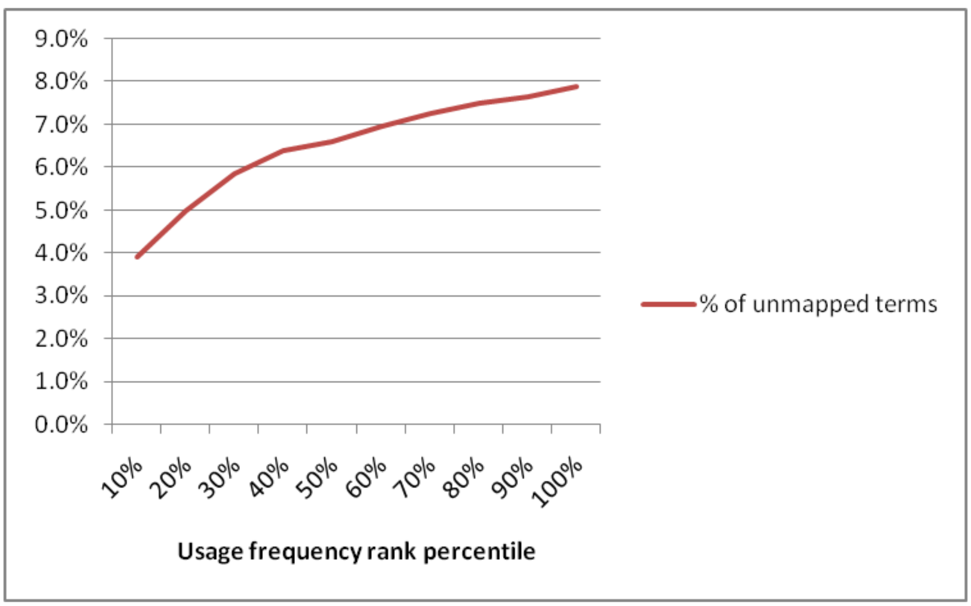


Figure 5. Relationship of term usage to mappability to the UMLS

Figure 5 shows the relationship between frequency of usage and mapping to the UMLS. The terms were listed in the descending order of their usage frequency. At each percentile point (e.g. 10^{th} percentile means 10% of the most frequently used terms) the percent of terms that could be mapped to the UMLS was plotted. The terms less used were also less mappable to the UMLS. This is not surprising because more commonly used terms are more likely to appear in standard terminologies and will be found in the UMLS.

5.2.4 Overlap between the datasets

Overlap between the local vocabularies was calculated based on the number of overlapping UMLS concepts (CUIs) across datasets. The pairwise overlap (Table 5) was calculated as follows:

$$\text{Percent of overlap of A with B} = \frac{\text{CUIs common to both A and B}}{\text{Number of CUIs in A}} \times 100\%$$

	Percent overlap with					
	HA	**IH**	**KP**	**MA**	**NU**	**RI**	**Average**
HA	-	17%	30%	34%	34%	13%	26%
IH	37%	-	66%	63%	76%	37%	56%
KP	29%	29%	-	51%	50%	19%	36%
MA	25%	21%	39%	-	46%	15%	29%
NU	27%	27%	41%	49%	-	20%	33%
RI	40%	50%	61%	62%	78%	-	58%

Table 5. Pairwise overlap between the datasets

The pairwise overlap showed high degree of variability among datasets. The lowest overlap was between HA with RI (13%) and the highest was between RI and NU (78%). HA had the lowest average overlap (26%) with other datasets while RI had the highest (58%). The overall average overlap for all datasets was 40%.

Total number of datasets that a CUI appeared in	**Number of CUIs**	**Percent of total**
1 dataset	4,201	62%
2 datasets	1,130	17%
3 datasets	607	9%
4 datasets	391	6%
5 datasets	282	4%
6 datasets	165	2%

Table 6. Distribution of CUIs among datasets

Another way to look at overlap was to see how often one CUI appeared in more than one dataset (Table 6). 4,201 concepts (62%) occurred in only one dataset. However, these 'unique' concepts are used much less frequently than the concepts that are common to at least two datasets. The 62% of unique concepts together only covered 15% of overall usage.

5.3 The CORE subset

It is not surprising that the local vocabularies only overlap to a modest degree. Each vocabulary is individually shaped by the unique characteristics of the health care institution and the way in which the problem list is used. These characteristics include: the disease epidemiology of the patient population served, distribution of the subspecialty service provided and the preferred level of granularity of the diagnostic terms. However, the observation that the terms that are shared among institutions are much more frequently used than the others is most interesting. These are the 38% of terms shared by at least 2 institutions which account for 77% of overall usage (compared to the other 62% of terms which only account for 15% of usage). This implies that it is possible to identify a core set of terms that are common to most problem list vocabularies which are used more frequently than others. That is the assumption on which the CORE subset idea is based.

The CORE subset can be used in two ways. For institutions that need to create a problem list vocabulary, the CORE subset can be used as a 'starter set'. As the local vocabulary is expanded by local extensions, it will inevitably diverge from the CORE subset. However, this divergence can be minimized if institutions use a standard way to add local extensions (see later discussion). A second way in which the CORE subset can be useful is for institutions to map their local problem list vocabularies to the CORE subset. This will facilitate the sharing of code problem list data across institutions. The goal of the CORE subset is to reduce the variability among coded problem list data and enhance data interoperability.

5.3.1 Candidates of the CORE subset

Ideally, the CORE subset should have the following features:

- High coverage of usage
- Small number of concepts
- Linkable to standard terminologies
- Supports reasoning
- Supports a standard mechanism to add local extensions

The CORE subset should cover a high percent of usage with a small number of terms. The ability of linking to standard terminologies (e.g. ICD9CM) will be a definite advantage. Some electronic medical record systems have additional intelligent functions (e.g. clinical decision support) that depend on the entries in the problem list. It will be useful if the problem list vocabulary has useful relationships (e.g. links between disease and body site, hierarchical structure of diseases) that can support electronic reasoning. A standard mechanism to add local extensions will minimize the tendency for local vocabularies to diverge.

Several approaches have been explored to arrive at the ideal CORE subset.

The starting point is to take everything. We shall call this the **UNION-SET**, which contains 6,776 concepts. From the UNION-SET, three subsets can be been derived as candidates for the CORE subset.

The **U-CORE** is the subset that covers 95% of the usage of the UNION-SET. This subset optimizes on coverage of usage. The U-CORE consists of 2,090 concepts.

The **O-CORE** is made up of concepts that exist in at least two datasets. All concepts that are unique to one dataset are excluded. The O-CORE consists of 2,575 concepts.

The **S-CORE** is the part of the UNION-SET that overlaps with SNOMED CT (only active SNOMED CT concepts are considered). Among the UMLS source terminologies, SNOMED CT is the single terminology that overlaps most with the UNION-SET (Table 7). The S-CORE consists of 5,348 concepts.

Terminology	Percent overlap
SNOMED CT	81%
MedDRA	64%
ICD9CM	49%
ICPC2 - ICD10 Thesaurus	41%
MeSH	31%
ICD10AM	28%

Table 7. Degree of overlap of the UNION-SET with individual terminologies

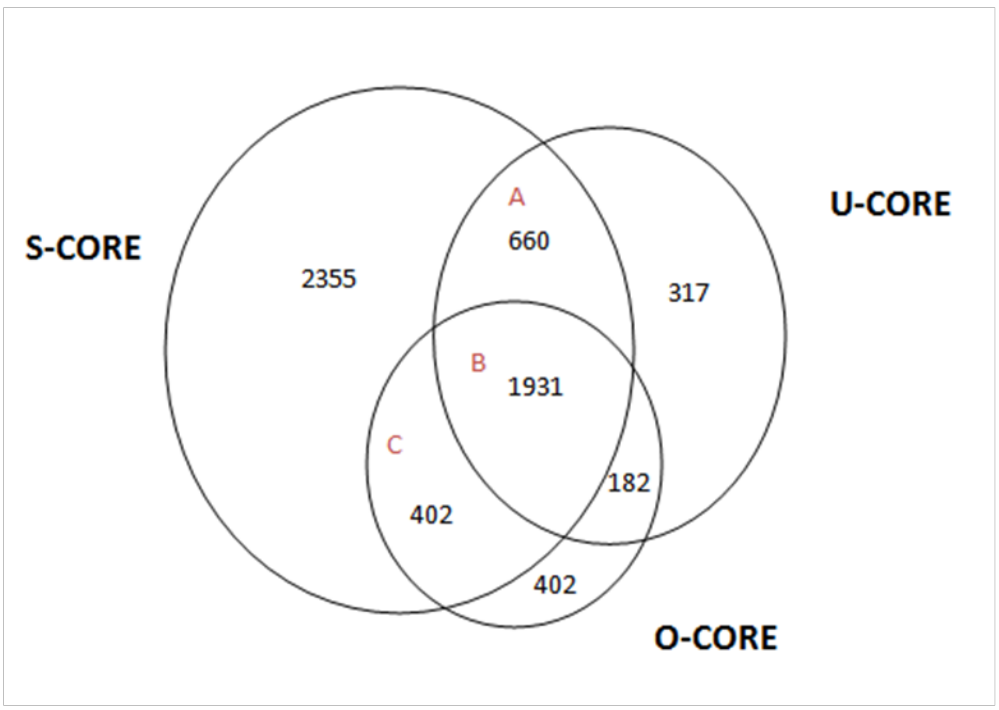


Figure 6. Overlap between S-CORE, U-CORE and O-CORE (I-CORE=A+B+C)

The three candidate CORE subsets overlap considerably with each other (Figure 6). There are 1,931 concepts (36% of S-CORE, 62% of U-CORE and 75% of O-CORE) that are common to all three subsets. Looking at their intersecting areas, another candidate CORE subset can be derived that combines the strong points of each of the three subsets. If we take the areas A + B + C, we will include the terms common to all three subsets, plus some high usage terms (area A) and some terms that occur in at least two datasets (area C). In addition, all the terms are SNOMED CT concepts (belonging to S-CORE). We shall call this the **I-CORE** and it contains 2,993 concepts.

5.3.2 Comparison of the candidate subsets

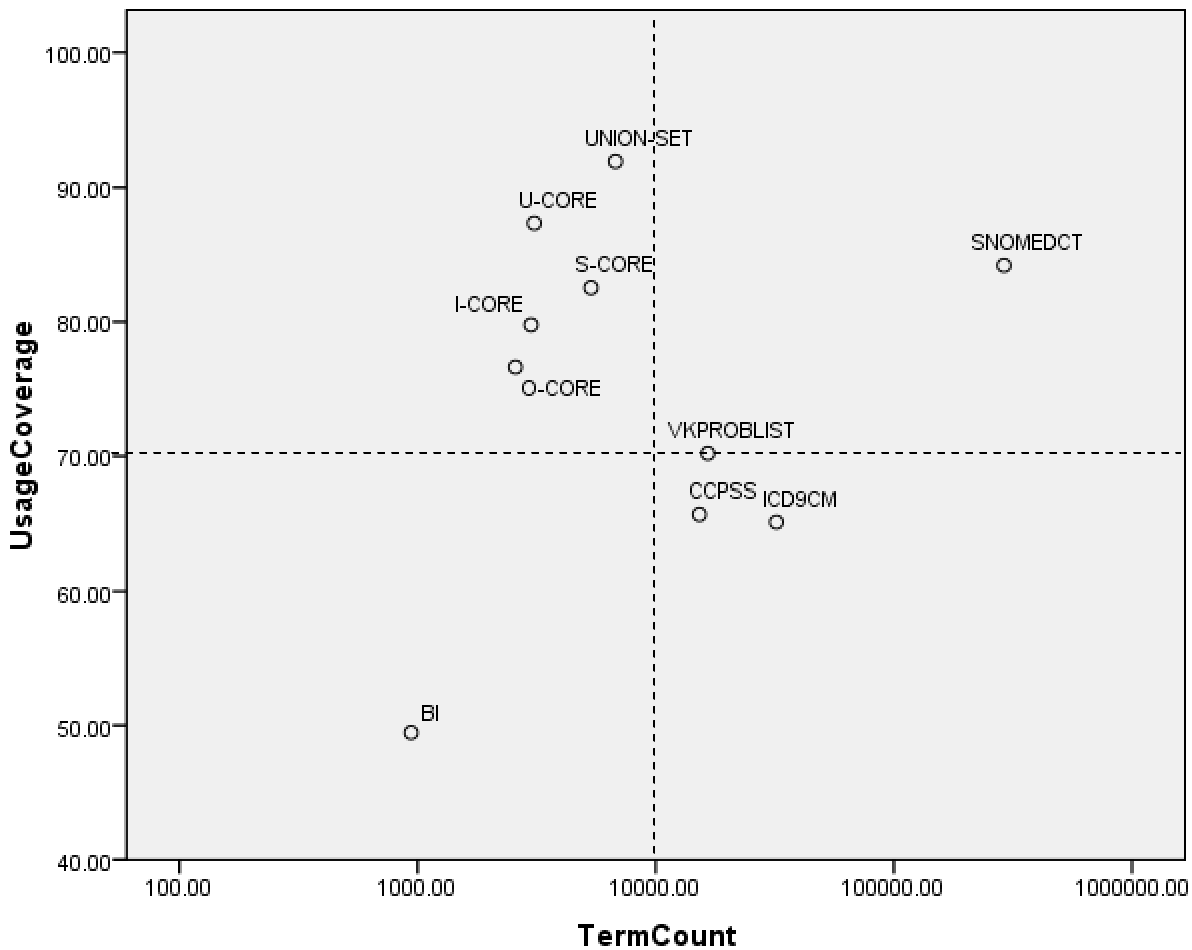


Figure 7. Usage coverage vs. size of the candidates CORE subsets and other terminologies

Figure 7 is a plot of usage coverage against the term count. The term count is expressed in logarithmic scale for better visualization, as the size of the terminologies spans over a wide range. Apart from the UNION-SET and the four candidate CORE subsets, other terminologies that can potentially be used as problem list vocabularies are also included for comparison. These terminologies are:

- SNOMED CT
- ICD9CM

- VA/KP Problem List Subset of SNOMED CT (VKPROBLIST)
- Canonical Clinical Problem Statement System (CCPSS)
- Beth Israel Clinical Problem List Vocabulary (BI)

The VA/KP Problem List Subset is one of the controlled terminologies used for Structured Product Label (SPL), a document markup standard approved by HL7 and adopted by FDA as a mechanism for exchanging medication information. It is developed by Kaiser Permanente and the VA, and is based on SNOMED CT. CCPSS and BI are problem list vocabularies developed by the Vanderbilt University and Beth Israel Deaconess Medical Center respectively and are published in the UMLS.

In Figure 7, the most desirable vocabularies are the ones closest to the left upper corner, i.e. high usage coverage and small size. The two dotted lines demarcate a limit of 10,000 concepts and usage coverage of 70%. Only the UNION-SET and the 4 candidate CORE subsets have less than 10,000 concepts while covering more than 70% of usage. One can further quantify the relationship between size and usage coverage by calculating the **coding density**, defined as the ratio of usage coverage to size. A higher number is more favorable. The O-CORE has the highest coding density (0.030) while the UNION-SET has the lowest (0.014) (Table 8).

Choosing a smaller CORE subset vocabulary can also reduce unnecessary variation in coding. There are terms in the UNION-SET which can be treated as almost synonymous in the clinical context. For example, there are 7 different terms for Gastroenteritis:

a. C0017160 Gastroenteritis
b. C0277525 Infectious gastroenteritis
c. C0267418 Non infectious gastroenteritis
d. C0318712 Viral gastroenteritis
e. C0277534 Gastroenteritis of presumed infectious origin
f. C1279224 Infectious colitis, enteritis, and gastroenteritis
g. C0029512 Noninfectious gastroenteritis and colitis

For most intents and purposes, only the first four (a - d) are clinically distinct concepts. The other three (e – f) do not represent clinically distinct entities. It is not uncommon that a diagnosis of infectious gastroenteritis is implied by the clinical finding without the support of microbiological test results, so b and e can be considered synonymous in most clinical settings. In a case of gastroenteritis, it is often not possible to ascertain whether the colon is actually involved by the infective process. After all, this is not an important distinction and will not alter patient management. So the distinction of b from f and c from g is not of clinical significance. If we can eliminate this redundancy, we can improve the overlap between the local vocabularies. In the above example, some of the quasi-synonymous terms are excluded if the smaller candidate CORE subsets are chosen: S-CORE has concepts a to f, U-CORE and I-CORE both have a to e and the smallest O-CORE has only a, c and d.

	UNION-SET	O-CORE	U-CORE	S-CORE	I-CORE
Size	6,776	2,575	3,090	5,348	2,993
Usage coverage	92%	77%	87%	83%	80%
Coding density	0.014	0.030	0.028	0.015	0.027
Linked to standard terminologies	Partial	Partial	Partial	A standard in itself, mapped to ICD9CM	A standard in itself, mapped to ICD9CM
Support reasoning	Limited	Limited	Limited	Yes	Yes
Mechanism for expansion	No	No	No	Yes	Yes

Table 8. Comparison of the candidate subsets

Table 8 summarizes the differences between the candidate subsets. The best choice in each category is highlighted in red. As discussed above, O-CORE is smallest and has the highest coding density while the UNION-SET has the highest usage coverage.

On the linkage to standard terminologies, all UMLS concepts can potentially be linked to standard terminologies through the UMLS concept structure. However, this is not 100% complete. For the UNION-SET, 81% of the concepts can be linked to SNOMED CT and 49% to ICD9CM (Table 7). On the other hand, all concepts in the S-CORE and I-CORE are already linked to SNOMED CT. In addition, through publicly available maps, most of these concepts can also be mapped to ICD9CM. The IHTSDO is currently also working on a map from SNOMED CT to ICD10.

Concerning the ability to support reasoning, SNOMED CT has a rich collection of well-defined relationships that can be used in electronic reasoning. To some extent, reasoning is also possible for UMLS concepts utilizing the relationships in the UMLS. However, the UMLS relationships are heterogeneous and they are not edited to ensure uniformity and consistency. The power and quality of the reasoning through UMLS relationships will not be as good compared to SNOMED CT relationships.

A core set of terms will not be enough to cover all the needs of every institution. There will always be the need to add local extensions. If this is not done in a disciplined way, local vocabulary lists will drift further and further apart. One advantage of the S-CORE and I-CORE is that there can be a standard way to add local extensions. To start with, one can look in the rest of SNOMED CT for additional terms. Failing that, the SNOMED CT rules for post-coordination can be used to represent new meaning by adding qualifiers (which are other SNOMED CT concepts) to the concepts in the CORE subset (e.g. adding a laterality qualifier). In this way, local problem lists will evolve in a more orderly fashion and most new terms will maintain their links to the CORE subset concepts.

5.4 Future work

More work is required to qualify the difference between the local problem list vocabularies. It will be interesting to know how much of that difference is due to genuine differences between institutions (e.g. disease epidemiology, subspecialty distribution) and how much is due to the presence of clinically quasi-synonymous terms. The use of a CORE subset can help to reduce the latter type of difference. The various approaches of defining the CORE subset need to be studied further to find the best candidate subset. Once defined, the CORE subset will be published for testing and comment.

6. Inter-terminology Mapping

6.1 The need for inter-terminology mapping

The need for mapping between medical terminologies commonly arises when data encoded in one terminology is reused for a secondary purpose that requires a different system of encoding. Imagine an electronic patient record system that captures clinical information using SNOMED CT codes. It will be a big efficiency gain if ICD9CM and CPT codes can be generated automatically for billing purposes. For this to happen, mapping from SNOMED CT to ICD9CM and CPT is required. Table 9 lists some possible use cases of inter-terminology mapping.

Primary purpose of data	Secondary use	Mapping requirement
Clinical documentation (encoded in SNOMED CT)	Service reimbursement (requires ICD9CM codes)	SNOMED CT => ICD9CM
Laboratory results reporting (encoded in LOINC)	Billing (requires CPT codes)	LOINC => CPT
Documentation of adverse drug reactions (encoded in SNOMED CT)	Reporting to regulatory institution (requires MedDRA codes)	SNOMED CT => MedDRA
Clinical problems list (encoded in SNOMED CT)	Literature search for decision support (requires MeSH codes)	SNOMED CT => MeSH

Table 9. Some use cases of inter-terminology mapping

Creating and maintaining maps is a labor-intensive process. The mapper needs to have in-depth knowledge of the both the source and target terminologies. A good interface to browse and search both terminologies is essential. In addition, mapping efficiency will be enhanced if the mapping tool can automatically suggest possible candidate maps with a high degree of accuracy. The focus of this research is to find algorithms that can identify candidate maps, making use of the resources in the UMLS. The UMLS is a particularly valuable resource in inter-terminology mapping because it contains a large number of source terminologies, a rich collection of relationships and the accompanying lexical tools.

6.2 Mapping algorithms explored

Generally speaking, inter-terminology mapping algorithms can be divided into lexically-based or semantically-based methods. Lexical methods rely on the lexical properties of the terms, which are usually normalized or broken down into segments before they are matched to the target terms. On the other hand, semantic methods find matches by utilizing semantic links between terms in the terminologies involved. This research has explored three different mapping algorithms: semantic, lexical and combined approaches.

6.2.1 Semantic mapping (IntraMap)

The IntraMap algorithm (a modification of the Restrict to MeSH algorithm) makes use of semantic relationships between UMLS concepts to find mappings. Starting from the source concept (the UMLS concept containing the term in a source terminology from which mapping is sought), the algorithm looks for target concepts (UMLS concepts containing terms in the terminology being mapped to) which are related to the source concept either through synonymy or explicit mapping relationships provided by some source terminologies. Failing to find a target concept, the search widens by using ancestors of the source concept as starting points to look for target concepts. If that fails again, ancestors of the children of the source concept, and finally, ancestors of the siblings of the source concept will be used for target concept searching.

6.2.2 Lexical mapping (MetaMap)

Lexical mapping make use of the MetaMap program which maps biomedical text to UMLS concepts. The name of the source concept is used as the input string to MetaMap and the output is limited to UMLS concepts containing terms from the target terminology. The MetaMap score is used to rank the output in terms of the likelihood of being a correct map.

6.2.3 Combined semantic and lexical mapping

Since semantic and lexical mapping are fundamentally different approaches, they are orthogonal and thus can be used to validate and complement each other. To make use of both sets of mappings simultaneously, a method was derived based on the precision level of each sub-category of mapping in the semantic and lexical algorithms. A precision ladder is constructed and the maps generated by the two methods are ranked according to the expected level of precision. In case where there are multiple alternative maps, only the one with the highest expected precision is kept.

6.2.4 Evaluation of the algorithms

Official maps from SNOMED CT to ICD9CM are periodically released by the IHTSDO. One of these maps (released in January 2004) was used as the gold standard. The version of UMLS used for the evaluation was 2004AA. All the SNOMED CT concepts having one-to-one maps in the gold standard were fed through the three algorithms to find candidate maps to ICD9CM. The candidate maps were compared to the gold standard and three metrics were calculated: coverage (percent of SNOMED CT terms for which candidate maps were found), recall (percent of maps in the gold standard that were found) and precision (percent of candidate maps that agreed with the gold standard).

6.3 Results

6.3.1 Semantic mapping alone

Among the 66,382 SNOMED CT terms, IntraMap managed to find ICD9CM maps for 57,293 terms (86.3% coverage). Overall recall was 43.3% and precision was 22.1%. On average, there were 2.3 maps found per SNOMED CT term. The precision of the sub-categories of mappings were: mapping by synonymy 78.4%, mapping by explicit mapping relationships 50.1% and mapping by ancestor expansion 9.2%. The mappings found by children and sibling expansion were too small in number to warrant further consideration. The results are summarized in Table 10.

	Sub-category of map			**Overall**
	Synonymy	**Explicit map relation**	**Ancestor expansion**	
Coverage	19.5%	17.0%	47.2%	86.3%
Recall	16.6%	13.0%	13.0%	43.3%
Precision	78.3%	50.1%	9.2%	22.1%
Map per term	1.1	1.5	3.0	2.3

Table 10. Overall and sub-category performance of semantic mapping by IntraMap

6.3.2 Lexical mapping alone

MetaMap was able to find maps for 44,452 SNOMED CT terms (70.0% coverage). The overall recall and precision was 28.4% and 14.7% respectively. There were on average 2.9 maps found per SNOMED CT term. Among those maps that were considered to be perfect matches (MetaMap score of 1000), the precision was 85.8%. For those SNOMED CT terms with no perfect matches found, if we only used the top ranking maps (the maps with the highest MetaMap score), the precision was 22.6%. The results are summarized in Table 11.

	Sub-category of map		**Overall**
	Perfect map	**Top map**	
Coverage	9.7%	57.2%	70.0%
Recall	8.6%	18.3%	28.4%
Precision	85.8%	22.6%	14.7%
Map per term	1.0	1.4	2.9

Table 11. Overall and sub-category performance of lexical mapping by MetaMap

6.3.3 Overlap between semantic and lexical mapping

A total of 29,468 mappings (distinct pairs of SNOMED CT and ICD9CM codes) were common to both sets of maps. This represented 22.6% and 22.9% of all IntraMap and MetaMap maps, respectively. This set of common maps covered 35.7% of the SNOMED CT terms, with recall and precision of 22.5% and 50.8% respectively. The maps that were only found in one algorithm but not the other were higher in coverage but lower in precision (Table 12).

	Both IntraMap and MetaMap	Only from IntraMap	Only from MetaMap
Coverage	35.7%	57.4%	51.9%
Recall	22.5%	20.8%	5.9%
Precision	50.8%	13.7%	3.9%
Map per term	1.2	2.6	2.9

Table 12. Mapping performance according to overlap between semantic and lexical mapping

Altogether 13,797 correct maps were found by semantic mapping alone and missed by lexical mapping. Among these, maps found by synonymy, explicit mapping relationship and ancestor expansion constituted 9%, 48% and 43% respectively. One example was the mapping from SNOMED CT term '3072001: Hormone-induced hypopituitarism' to ICD9CM term '253.7: Iatrogenic pituitary disorders'. The failure of MetaMap to find this mapping was expected as it was unlikely that the similarity in meaning between 'hormone-induced' and 'iatrogenic' could be detected by lexical matching alone.

On the other hand, there were 3,906 correct maps that were found by lexical mapping but missed by semantic mapping. Most of these were maps from a narrower to a broader concept. One example was the map from the SNOMED CT term '67600007: Vascular-biliary fistula' to the ICD9CM term '576.4: Fistula of bile duct' by way of the synonym 'biliary fistula' in the same UMLS concept. This map was not found by IntraMap because the two UMLS concepts containing the two terms were not linked by any hierarchical or mapping relationships in the UMLS.

6.3.4 Combined semantic and lexical mapping

A precision ladder was created based on the level of precision of the subcategories of each mapping approach and whether a map was found in both sets. As shown in Table 13, the highest ranking subcategory was the MetaMap perfect matches and the lowest was those maps that were present only in MetaMap.

Rank	Sub-category	Precision
1	M-PM (MetaMap perfect match)	85.8%
2	I-S (IntraMap synonymy)	78.3%
3	C-O (Combined overlapping)	50.8%
4	I-EM (IntraMap explicit mapping)	50.1%
5	M-TM (MetaMap top score)	22.6%
6	C-IO (Combined IntraMap only)	13.7%
7	I-AE (IntraMap ancestor expansion)	9.2%
8	C-MO (Combined MetaMap only)	3.9%

Table 13. Precision ladder according to precision of each sub-category of mapping

All the maps were pooled together and arranged in descending order of expected precision according to the precision ladder. If the same map appeared in more than one sub-category, only the one in the highest ranking sub-category was kept. If there were multiple mappings for the same SNOMED CT term, only the one with the highest ranking was kept and the alternative lower ranking maps were discarded. The combined set contained 107,172 maps for 60,454 SNOMED CT terms (coverage 91.1%). The overall recall and precision of the combined set was 42.9% and 26.6% respectively.

The fact that the maps were arranged in the order of precision provided a further means of fine-tuning the performance according to the way that the maps were intended to be used. By setting different cut-off points on the precision ladder (i.e. ignoring mappings below a certain rank) one could obtain different combinations of coverage, recall and precision. As expected, the further down the precision ladder, the higher the coverage and recall but the lower the precision (Table 14). The mapping performance did not change further with inclusion of maps ranking lower than rank 6. This was because the maps in rank 7 (IntraMap maps found by ancestor expansion) were already included in higher sub-categories (ranks 3 and 6). Maps from rank 8 did not contribute to the overall performance because rank 1 and rank 5 already covered every SNOMED CT term for which a map was found by MetaMap. As a single indicator of performance, the F-score with equal emphasis on recall and precision (F-score = $(0.5/\text{precision} + 0.5/\text{recall})^{-1}$) was calculated for each cut-off point. Judging from the F-score, it seems that rank 4 is the optimal cut-off point if recall and precision are equally important.

	Cut-off point on the precision ladder							
	Rank 1	**Rank 2**	**Rank 3**	**Rank 4**	**Rank 5**	**Rank 6**	**Rank 7**	**Rank 8**
Coverage	9.7%	19.5%	38.1%	50.8%	71.0%	91.1%	91.1%	91.1%
Recall	8.6%	16.6%	24.4%	34.0%	38.2%	42.9%	42.9%	42.9%
Precision	85.8%	78.4%	52.0%	51.6%	39.3%	26.6%	26.6%	26.6%
Map per term	1.0	1.1	1.2	1.3	1.4	1.8	1.8	1.8
F-score	0.16	0.27	0.33	0.41	0.39	0.33	0.33	0.33

Table 14. Mapping performance according to the cut-off point on the precision ladder

The advantage of the combined method is twofold. Firstly, the combined method performs better than either the semantic or lexical mapping alone. The overall coverage, recall and precision of the combined method were 91%, 43% and 27% respectively, which were better than the semantic (86%, 43%, 22%) or lexical (70%, 28%, 15%) method used individually. Secondly, the precision ladder provides the possibility of adjusting the recall-precision profile of the candidate maps to suit the task at hand. For instance, if the task is automatic code translation, one would generally prefer a mapping algorithm with high precision. One way to achieve this is by taking only the first two rungs of the precision ladder. This will give highly precise mappings (precision close to 80%) to about 20% of SNOMED CT terms. On the other hand, a more likely use case of the mapping algorithms is to suggest candidate maps to human editors creating a map. In this situation, one will prefer a mapping algorithm with high coverage. If one takes every rung from the precision ladder, one will have candidate maps for over 90% of SNOMED CT terms, with a precision of 27%. The precision may seem a bit low, but in this use case, even the incorrect candidate maps may serve some useful purpose. If we consider only the first three digits of the ICD9CM codes, the precision jumps to almost 50%. This means that one out of two of the candidate maps will either be exactly correct or will bring the editors closer to the correct map.

6.4 Inter-terminology mapping in practice

Since SNOMED CT became available to all U.S. users and its designation as the U.S. standard for clinical documentation, there has been an ongoing demand for maps from SNOMED CT to other terminologies. It is conceivable that such maps will help to promote the use of SNOMED CT in electronic patient records. NLM has initiated, facilitated or directly funded several mapping projects:

- SNOMED CT to ICD9CM rule-based map for reimbursement
- SNOMED CT to CPT
- LOINC to CPT
- SNOMED CT to MeSH

The early versions of some of these maps have already been released for public testing and comments. At the international level, there is a special project group under IHTSDO to map SNOMED CT to ICD10.

The above research on automatic mapping serves as proof-of-concept of methods to use the UMLS to map between terminologies. Even though no ready-to-use tools have been developed, the automatic mapping algorithms could be used in any of these mapping projects to enhance the efficiency of human mapping.

7. The RxHub Project – a sneak preview

A brief account of a still ongoing project, the RxHub Project, is included here as an example of a research project that involves the use of medical terminologies. This is intended only as a 'sneak preview'. A formal report will be presented when the project is finished.

7.1 Background

Physicians consistently report that information about current and past medication usage is one of the most important categories of information for emergency patient care. This information is important for deciding what could be causing the patient's current medical problem, for choosing the medication used to treat that problem, for inferring the existence of concurrent diagnoses, and for avoiding dangerous interactions (e.g. between Monoamine Oxidase inhibitors and Tricyclic antidepressants). Most care systems/providers do collect a medication history, but the intensity and timeliness of this process vary greatly. The Joint Commission on Accreditation of Healthcare Organizations requires that hospitals do medication reconciliation on all patients who are admitted.

SureScripts-RxHub is a consortium of major pharmacy benefit management systems covering 60-70% of the prescriptions paid for by private insurance in the U.S. The insurance claims information which contains the name of the prescribed drugs, dosage and duration is a valuable source of medication information. SureScripts-RxHub offers a service that provides this information to health care institutions for patient care. This can be potentially very useful in enhancing the accuracy and efficiency of medication reconciliation.

This project is funded by the Bethesda Hospitals' Emergency Preparedness Partnership (BHEPP) a collaboration formed by three healthcare facilities in the Bethesda area, namely National Naval Medical Center, NIH Clinical Center and Suburban Hospital. The overall goal of this partnership is to develop integrated mechanisms for providing immediate care to patients who might be harmed in a Washington area disaster and to integrate and organize the resources of the three facilities in order to provide an effective and sustained response to any local, regional or national emergency. This particular project in support of the partnership's goals is to develop mechanisms for providing prescription medication histories to clinicians caring for disaster patients.

7.2 Objectives and methods

The objectives of this project are twofold. Firstly, we want to evaluate the usefulness of SureScripts-RxHub prescription information in patient care. Secondly, we want to develop a predictive model of the availability of SureScripts-RxHub data based on demographics, insurance coverage and other patient characteristics.

We will establish a connection for Suburban Hospital to access SureScripts-RxHub data through the use of HL7 messages. SureScripts-RxHub requires 5 pieces of information: first name, last name, date of birth, gender and zip code in order to match to patients in their database. For all patients attending the Emergency Department, this information will be obtained from hospital initiated ADT messages (a type of HL7 message) and sent to SureScripts-RxHub. The returning SureScripts-RxHub message will fall into one of the following groups:

- No matching patient in SureScripts-RxHub database
- Patient found but no medication history
- Patient found with associated medication history

In the study period, the SureScripts-RxHub drug information will not be available to the medical staff to avoid contamination. During the same time, the medication history taken by the admitting staff, together with other relevant patient information will be retrieved from the hospital information system. The SureScripts-RxHub drug information will be compared to the manually-recorded medication information in terms of its accuracy and completeness. All data are properly de-identified before they are sent to us for analysis.

To be able to compare the two sources of medication information, we need to map them to a common drug terminology. This is where medical terminology is involved. We shall map the medication data to names in RxNorm/RxTerms. Mapping the manual medication history will not be easy because the information is entered as free-text. A combination of lexical matching with selective manual validation will be used. Mapping of the SureScripts-RxHub data will be easier as the data is structured and often accompanied by NDC codes.

7.3 Current status

We have installed the infrastructure necessary for Suburban Hospital to obtain SureScripts-RxHub data. End-to-end dataflow has begun since March. We expect the data collection to span over two months, after which SureScripts-RxHub drug data will be routinely available for patient care. We expect to collect data on about 7,000 patients. The results of this study will be reported separately when they are available.

8. Conclusion

This report highlights some of the research that I have done on medical terminologies. Medical terminologies and the UMLS project has been the main focus of my work since I joined NLM in 2003. Because of my clinical background and experience in building hospital information systems, I am inclined to formulate my research from a practical perspective. My research program is called Applied Medical Terminology Research to emphasize its focus on the practical issues and problems rather than more theoretical aspects of terminologies.

Through my research I hope that I can:

- make standard terminologies easier to use
- maximize the benefits of using them

RxTerms makes it easier to encode medication information with RxNorm. The CORE subset will provide a ready solution for users in need of a standard problem list vocabulary. It will help users to reap the benefits of data sharing. Inter-terminology mapping promotes the use of standard terminologies with the benefits of data re-use.

Despites decades of work, standardization of information representation through the use of controlled terminologies remains one of the biggest challenges in informatics research. However, without solving this problem, many of the benefits and promises of Medical Informatics cannot be realized. It is hoped that my research will contribute towards this goal.

9. Acknowledgements

I would like to express my sincere thanks to the following people who have contributed to the above projects.

RxTerms
Dr. Clement McDonald
Paul Lynch
Kevin Lai

Problem list vocabularies
Dr. Clement McDonald
Dr. Stan Huff
Dr. Chris Chute
Dr. James Campbell
Dr. Robert Dolin
Vicky Fung
Suresh Srinivasan
Dr. John Kilbourne

Inter-terminology mapping
Dr. Olivier Bodenreider
Dr. Alan Aronson

Special thanks to Dr William Hole (retired) for his kind support and guidance. Also thanks to Dwayne McCully who maintains our terminology research database.

10. References

1. Bates DW, Gawande AA. Improving safety with information technology. N Engl J Med. 2003 Jun 19;348(25):2526-34.

2. Jha AK, Ferris TG, Donelan K, et al. How common are electronic health records in the United States? A summary of the evidence. Health Aff (Millwood). 2006 Nov-Dec;25(6):w496-507.

3. Saathoff A. Human factors considerations relevant to CPOE implementations. J Healthc Inf Manag. 2005 Summer;19(3):71-8.

4. Rosenbloom ST, Miller RA, Johnson KB, Elkin PL, Brown SH. Interface terminologies: facilitating direct entry of clinical data into electronic health record systems. J Am Med Inform Assoc. 2006 May-Jun;13(3):277-88.

5. RxNorm documentation, NLM website
http://www.nlm.nih.gov/research/umls/rxnorm/index.html
Accessed Mar 2009.

6. Consolidated Health Informatics (CHI) Initiative; Health Care and Vocabulary Standards for Use in Federal Health Information Technology Systems. Federal Register. Dec, 2005;70, No.246.

7. Name Differentiation Project, FDA
http://www.fda.gov/CDER/Drug/MedErrors/nameDiff.htm
Accessed Mar 2009.

8. Most commonly prescribed drugs, EMC/Paradigm Publishing College Resource Centers
http://www.emcp.com/college_resource_centers/index.php?GroupID=7237
Accessed Mar 2009.

9. Fung KW, McDonald C, Bray B. RxTerms – A Drug Interface Terminology Derived from RxNorm. AMIA Annu Symp Proc 2008: 227-231

10. Rector A. Clinical Terminology: Why is it so hard? Method Inform Med 1999; 38:239-52

11. Brown SH, Miller RA, Camp HN, Guise DA, Walker HK. Empirical derivation of an electronic clinically useful problem statement system. Ann Intern Med 1999;131:117-126

12. Humphreys BL, McCray AT, Cheh ML. Evaluating the coverage of controlled health data terminologies: Report on the results of the NLM/AHCPR Large Scale Vocabulary Test. J Am Med Inform Assoc 1997;4:484-500

13. Penz JFE, Brown SB, Carter JS, Elkin PL, Nguyen VN, Sims SA, Lincoln MJ. Evaluation of SNOMED coverage of Veterans Health Administration terms. Medinfo 2004;107(Pt 1):540-4

14. Travers DA, Haas SW. Unified medical language system coverage of emergency-medicine chief complaints. Acad Emerg Med. 2006 Dec;13(12):1319-23

15. Rosenbloom ST, Brown SH, Froehling D, Bauer BA, Wahner-Roedler DL, Gregg WM, Elkin PL. Using SNOMED CT to represent two interface terminologies. J Am Med Inform Assoc. 2009 Jan-Feb;16(1):81-8

16. Elkin PL, Brown SH, Husser CS, Bauer BA, Wahner-Roedler D, Rosenbloom ST, Speroff T. Evaluation of the content coverage of SNOMED CT: ability of SNOMED clinical terms to represent clinical problem lists. Mayo Clin Proc. 2006 Jun;81(6):741-8

17. Brown SH, Elkin PL, Bauer BA, Wahner-Roedler D, Husser CS, Temesgen Z, Hardenbrook SP, Fielstein EM, Rosenbloom ST. SNOMED CT: utility for a general medical evaluation template. AMIA Annu Symp Proc. 2006:101-5

18. Klimczak JC, Hahn AW, Sievert ME, Petroski G, Hewett J. Comparing clinical vocabularies using coding system fidelity. Proc Annu Symp Comput Appl Med Care. 1995:883-7

19. Lorence DP, Ibrahim IA. Disparity in coding concordance: do physicians and coders agree? J Health Care Finance. 2003 Summer;29(4):43-53

20. Lorence D. Regional variation in medical classification agreement: benchmarking the coding gap. J Med Syst. 2003 Oct;27(5):435-43

21. Yao P, Wiggs BR, Gregor C, Sigurnjak R, Dodek P. Discordance between physicians and coders in assignment of diagnoses. Int J Qual Health Care. 1999 Apr;11(2):147-53

22. Letrilliart L, Guiguet M, Flahault A. Reliability of report coding of hospital referrals in primary care versus practice-based coding. Eur J Epidemiol. 2000;16(7):653-9

23. Chiang MF, Hwang JC, Yu AC, Casper DS, Cimino JJ, Starren JB. Reliability of SNOMED-CT coding by three physicians using two terminology browsers. AMIA Annu Symp Proc. 2006:131-5

24. Hwang JC, Yu AC, Casper DS, Starren J, Cimino JJ, Chiang MF. Representation of ophthalmology concepts by electronic systems: intercoder agreement among physicians using controlled terminologies. Ophthalmology. 2006 Apr;113(4):511-9

25. Fung KW, Hole WT, Nelson SJ, Srinivasan S, Powell T, Roth L. Integrating SNOMED CT into the UMLS: an exploration of different views of synonymy and quality of editing. J Am Med Inform Assoc. 2005 Jul-Aug;12(4):486-94

26. Fung KW, Bodenreider O, Aronson AR, Hole WT, Srinivasan S. Combining lexical and semantic methods of inter-terminology mapping using the UMLS. Medinfo 2007:605-609

27. Fung KW, Bodenreider O. Utilizing the UMLS for semantic mapping between terminologies. AMIA Annu Symp Proc. 2005:266-70

28. Rocha RA, Rocha BH, Huff SM. Automated translation between medical vocabularies using a frame-based interlingua. Proc Annu Symp Comput Appl Med Care 1993:690-4

29. Cimino JJ, Barnett GO. Automated translation between medical terminologies using semantic definitions. MD Comput 1990;7:104-9

30. Barrows RC, Jr., Cimino JJ, Clayton PD. Mapping clin-ically useful terminology to a controlled medical voca-bulary. Proceedings - the Annual Symposium on Com-puter Applications in Medical Care 1994:211-5

31. Dolin RH, Huff SM, Rocha RA, Spackman KA, Camp-bell KE. Evaluation of a "lexically assign, logically refine" strategy for semi-automated integration of overlapping terminologies. Journal of the American Medical Informatics Association 1998;5:203-13

32. Masarie FE, Jr., Miller RA, Bouhaddou O, Giuse NB, Warner HR. An interlingua for electronic interchange of medical information: using frames to map between clinical vocabularies. Comput Biomed Res 1991;24:379-400

33. Rocha RA, Huff SM. Using digrams to map controlled medical vocabularies. Proc Annu Symp Comput Appl Med Care 1994:172-6

34. Bodenreider O, Nelson SJ, Hole WT, Chang HF. Beyond synonymy: exploiting the UMLS semantics in mapping vocabularies. Proceedings / AMIA Annual Symposium 1998:815-9

35. Aronson AR. Effective mapping of biomedical text to the UMLS Metathesaurus: the MetaMap program. Proceedings / AMIA Annual Symposium 2001:17-21

36. United States General Accounting Office. Automated medical records: Leadership needed to expedite standards development. USGAO-IMTEC-93-17. Washington D.C., 1993

37. Cimino JJ. Review paper: coding systems in health care. Methods of Information in Medicine 1996;35:273-84

38. Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. Methods of Information in Medicine 1993;32:281-91

39. Humphreys BL, Lindberg DA, Schoolman HM, Barnett GO. The Unified Medical Language System: an informatics research collaboration. J Am Med Inform Assoc 1998;5:1-11

40. Bodenreider O. Circular hierarchical relationships in the UMLS: etiology, diagnosis, treatment, complications and prevention. Proceedings / AMIA . Annual Symposium 2001:57-61

11. Questions for the Board

1. What other features would make RxTerms more useful? Things that we are considering include drug class information (derived from NDF-RT) and links to on-line drug information sources (e.g. DailyMed).

2. Among the candidate CORE subsets, which is the most appropriate for its purpose? Are there other candidates that we should consider? Should we consider publishing more than one CORE subset?

3. How much impact will the CORE subset have in reducing the variability of problem list vocabularies? What can we do to promote the use of the subset?

Curriculum Vitae

Kin Wah Fung

POSITIONS

2005 - present	Staff scientist, Lister Hill National Center for Biomedical Communications, NLM, NIH
2003 – 2005	Postdoctoral Fellow, Lister Hill National Center for Biomedical Communications, NLM, NIH
1999-2001	Senior Executive Manager in Health Informatics, Hong Kong Hospital Authority
1990-1999	Consultant surgeon, Kwong Wah Hospital, Hong Kong
1997-1999	Chief coordinator for the development of clinical information systems in Kwong Wah Hospital, Hong Kong
1993-1999	Honorary assistant clinical professor, Department of Surgery, University of Hong Kong
1984-1990	Surgical resident, Queen Elizabeth Hospital, Hong Kong

EDUCATION AND QUALIFICATIONS

2002-2003	Master of Arts in Medical Informatics, Columbia University, New York
1994-1997	Master of Science in Computer-based Information Systems, University of Sunderland, UK (graduated with Distinction)
1993	Fellow of Hong Kong Academy of Medicine (Plastic Surgery Board)
1988	Fellow of Royal College of Surgeons of Edinburgh, UK
1979-1984	Bachelor of Medicine and Bachelor of Surgery, University of Hong Kong

AWARDS

National Library of Medicine Special Act or Service Award, 2008 – for the development of RxTerms, which provides a free, user-friendly and efficient drug interface terminology linked to RxNorm

National Library of Medicine Special Act or Service Award, 2004 – in recognition of the contribution to the successful integration of SNOMED CT into the UMLS

PUBLICATIONS

1. Fung KW, McDonald C, Bray B. RxTerms – A Drug Interface Terminology Derived from RxNorm. AMIA Annu Symp Proc 2008: 227-231

2. Sneiderman C, Demner-Fushman D, Fung KW, Bray B. UMLS-based Automatic Image Indexing. AMIA Annu Symp Proc 2006: 1141

3. Richesson RL, Fung KW, Krischer JP. Heterogeneous but "Standard" Coding Systems for Adverse Events: Issues in Achieving Interoperability between Apples and Oranges. Contemporary Clinical Trials 2008 Sep;29(5):635-45.

4. Fung KW, Bodenreider O, Aronson AR, Hole WT, Srinivasan S. Combining lexical and semantic methods of inter-terminology mapping using the UMLS. Medinfo 2007:605-609. (Finalist for Best Paper Award)

5. AR Aronson, Bodenreider O, Demner-Fushman D, Fung KW, Lee VK, Mork JG, Neveol A, Peters L, Rogers WJ. From Indexing the Biomedical Literature to Coding Clinical Text: Experience with MTI and Machine Learning Approaches. Proc BioNLP 2007 Workshop, 105-12.

6. Fung KW, Hole WT, Srinivasan S. Who is Using the UMLS and How – Insights from the UMLS User Annual Reports. AMIA Annu Symp Proc 2006: 274-278

7. Fung KW, Bodenreider O. Utilizing the UMLS for Semantic Mapping between Terminologies. AMIA Annu Symp Proc 2005:266-70

8. Fung KW, Hole WT, Nelson SJ, Srinivasan S, Powell T, Roth L. Integrating SNOMED CT into the UMLS: An Exploration of Different Views of Synonymy and Quality of Editing. Journal of American Informatics Assoc. 2005; 12(4): 486-94

9. Fung KW, Vogel LH. Will decision support in medications order entry save money? A return on investment analysis of the case of the Hong Kong Hospital Authority. AMIA Annu Symp Proc 2003:244-54

10. Cheung NT, Fung KW, Wong KC et al. Medical Informatics – the state of the art in the Hospital Authority. International Journal of Medical Informatics 2001;62(2-3): 113-9

11. Fung V, Cheung NT, Chu YC, Chow YY, Tung Y, Wong KC, Fung KW et al. Statistical data collection as a by-product of clinical documentation. APAMI – MIC 2000 Conference, Hong Kong, Sept 2000.

12. Kwok CK, Cheung NT, Fung KW et al. The electronic patient record in the Hospital Authority. APAMI – MIC 2000 Conference, Hong Kong, Sept 2000

13. Fung KW, Lau Y, Fielding R, Or A and Yip AWC. The impact of mastectomy, breast-conserving treatment and immediate breast reconstruction on the quality of life of Chinese women. Australia and New Zealand Journal of Surgery 2001; 71:202-6

14. Fung KW and Yip AWC. Mortality Audit Based on a Model of Potentially Avoidable Adverse Events. Annual Surgical Audit Conference, Hospital Authority, Hong Kong, Jan 1999

15. Fung KW, Lau Y, Nyunt K, Chan CM and Ho CM. Immediate breast reconstruction in Chinese women using the transverse rectus abdominis myocutaneous (TRAM) flap. Australia and New Zealand Journal of Surgery,1996; 66: 452-6

16. Fung KW and Yip AWC. Tuberculous Splenic Abscess: Diagnosis by Fine Needle Aspiration Cytology. Asian Journal of Surgery, 1995;18(2):142-144