Specific Use Cases of SNOMED CT in the Electronic Health Record

Kin Wah Fung, MD, MS, MA

Lister Hill National Center for Biomedical Communications
National Library of Medicine, NIH

Outline

- To encode the problem list (or other summary sections) of the EHR
- Generation of administrative codes through mapping

CORE Problem List Subset of SNOMED CT

The problem list

- Problem-oriented medical record (Weed 1971) – not widely adopted
- The problem list – a complete list of the patient's problems, including diagnosis, findings or symptoms has become an essential feature of most paper and electronic health records (EHR)
- In an EHR, the problem list is often the first (if not the only) place to find encoded clinical data
- Incentives to codify problem list data:
  - Automatic generation of codes for billing and reporting
  - Clinical decision support
  - Implementation of care plan
  - Enrollment to research protocol

The UMLS-CORE Project

- Started in 2007
- UMLS (Unified Medical Language System) – a flagship terminology project of NLM
- CORE (Clinical Observations Recording and Encoding) – the capture and codification of clinical information in the summary segments of the EHR such as the problem list, discharge diagnosis and reason for encounter
- To study the problem list terminologies of large health care institutions and identify a subset of concepts that are used most frequently

Methods

- Large health care institutions are asked to submit their problem list vocabularies, together with their actual frequency of use
- Local terms are mapped to the UMLS
- Analyze overlap
- Derive ways of defining the CORE subset

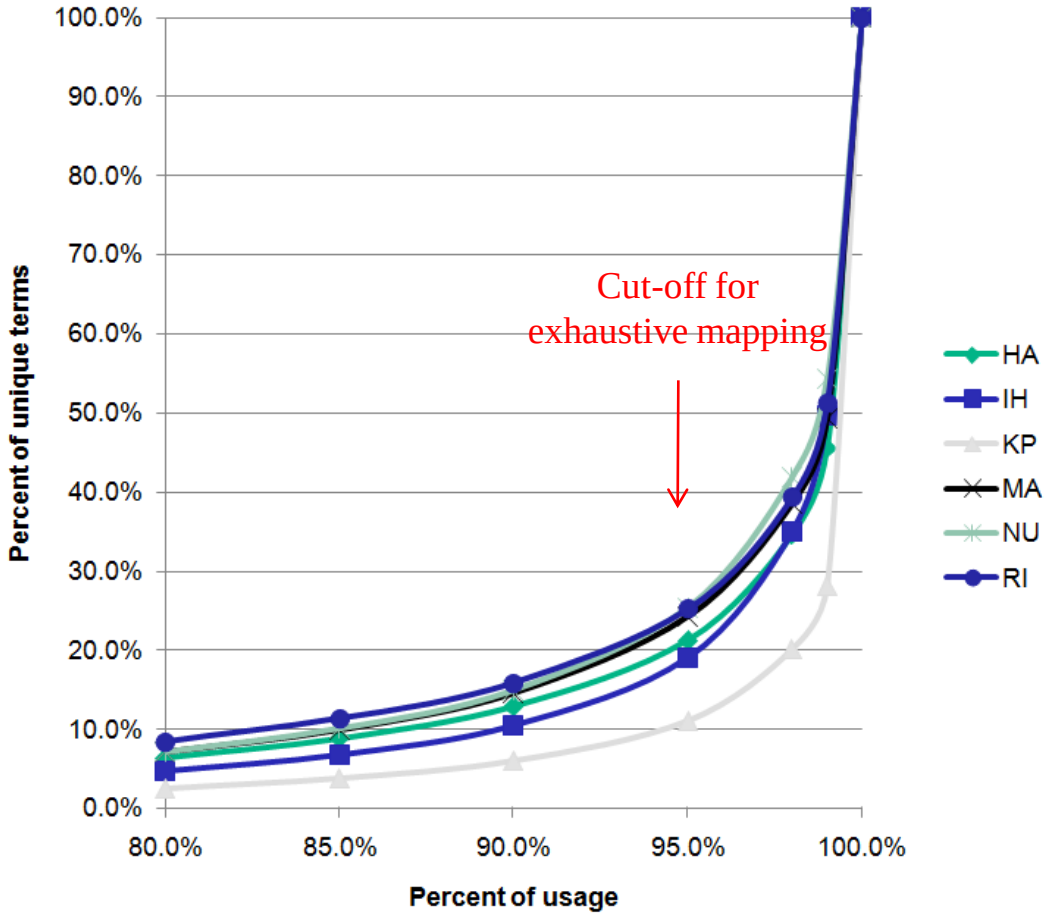
Data sets

- Problem list terms and usage frequencies are collected from 7 large health care institutions
  - Hong Kong Hospital Authority
  - Intermountain Healthcare
  - Kaiser Permanente
  - Mayo Clinic
  - Nebraska University Medical Center
  - Regenstrief Institute
  - Beth Israel Deaconess Medical Center (acquired much later, not included in data analysis)
- Data cover 14 million patients from all major medical specialties

The data sources

	HA	IH	KP	MA	NU	RI
Type of service	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
Size of vocabulary	12,449	5,685	26,890	14,921	13,126	3,166

Usage pattern



20% of terms cover about 95% of usage

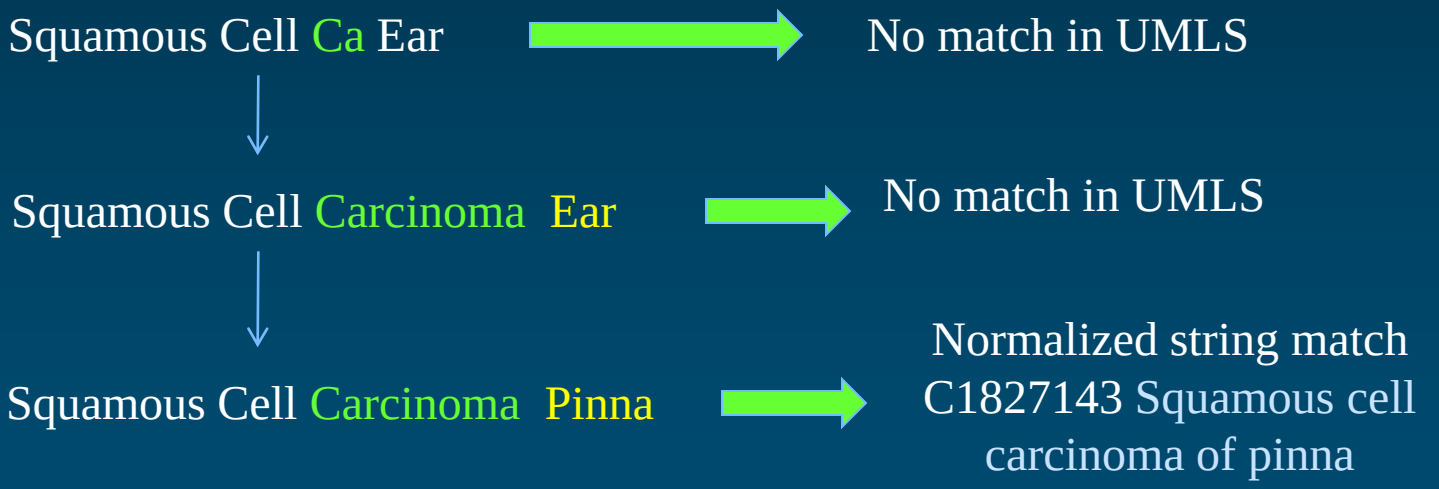
long tail of very infrequently used terms

Mapping to UMLS concepts

- Altogether over 70,000 local terms
- Only highest ranking 14,395 terms covering 95% usage in each set were mapped
- Mapping done in 3 steps:
  - Lexical mapping
  - Mapping via local maps
  - Manual mapping

Lexical mapping

- Exact, case insensitive and normalized string matches to all English strings in the UMLS
- Re-matching after synonymous word substitution (up to 2 substitutions)



Lexical mapping results

- Overall 75% of local terms can be mapped lexically:
  - Exact match 46%
  - Normalized string 14%
  - Case insensitive 10%
  - Synonym substitution 5%

Mapping by local maps

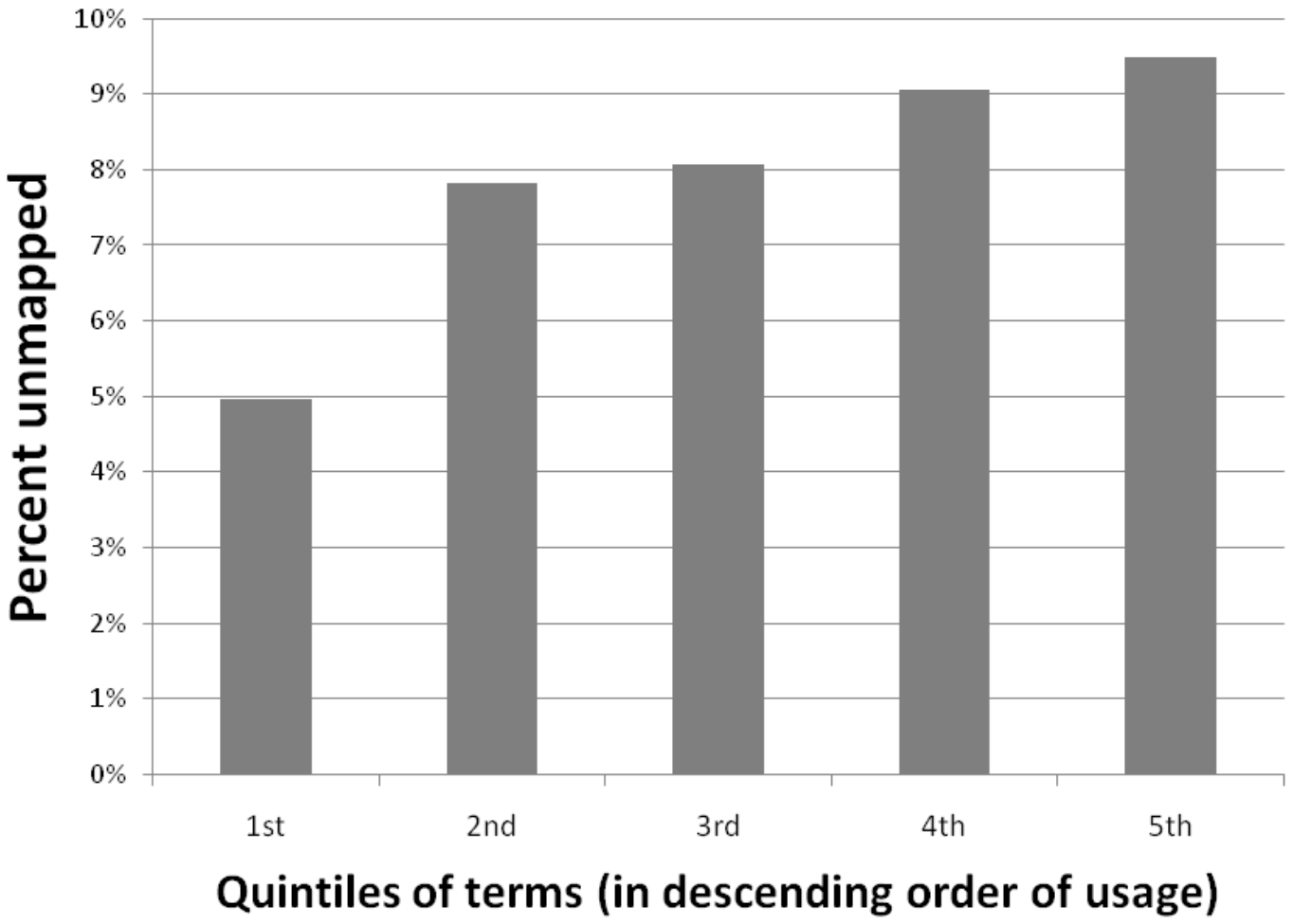
- Some datasets came with local maps to ICD9CM or SNOMED CT
  - Maps explicitly labeled as synonymous maps – used to map to the UMLS
  - All other maps are manually validated
- Local maps yielded UMLS maps to a further 7% of terms

Manual mapping

- Remaining 2,576 terms mapped by manual mapping, using RRF browser as search tool
  - yielded map for a further 10% of terms
- Overall 1,134 (8%) terms cannot be mapped to the UMLS
- Distribution of unmappable terms (%):

	HA	IH	KP	MA	NU	RI
% Not mapped	10%	2%	9%	11%	4%	4%

Less used terms are less mappable



Categories of terms not found in standard terminologies

Category	%	Example
Highly specific	53%	Benign prostatic hyperplasia with age related prostate cancer risk and obstruction
Very general	11%	Abnormal blood finding
Administrative	7%	Other Mr # exists
Laterality	7%	Renal stone, right
Negative finding	3%	No urethral stricture
Composite concept	3%	Diarrhea with dehydration
Meaning unclear	2%	Conjunctiva Red
Miscellaneous	13%	Subacute sinusitis

Pairwise overlap by percent of terms

- Overlap was assessed based on the terms that could be mapped to the UMLS

$$\%\ \text{overlap of A and B} = \frac{\#\ \text{CUI in both A \& B}}{\#\ \text{CUI in either A or B}}$$

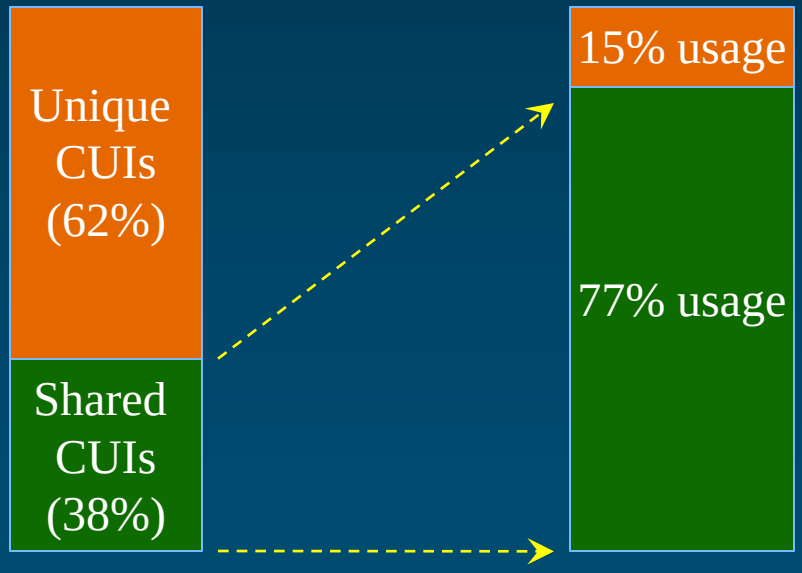
*CUI: Concept Unique Identifiers in the UMLS

Institution	Pairwise overlap with					
	IH	KP	MA	NU	RI	Mean pairwise overlap
HA	13%	17%	17%	18%	11%	15%
IH		25%	19%	25%	27%	22%
KP			29%	29%	17%	23%
MA				31%	14%	22%
NU					19%	24%
RI						18%

Distribution of CUIs among datasets

- CUI Distribution of CUIs among datasets:

# of datasets	1	2	3	4	5	6
% of CUIs	62%	17%	9%	6%	4%	2%



The terms that are shared among problem lists are used 8 times more frequently than those unique to one dataset => basis of the CORE subset

Desirable features of the CORE subset

- High coverage of usage by relatively small number of concepts
- Use of a wide-adopted terminology standard
- Supports reasoning
- Supports a standard mechanism for adding local extensions

Coverage of problem list concepts by standard terminologies

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

Problems of using SNOMED CT

- Institutions often only use a fraction of SNOMED CT concepts in their problem list vocabularies (e.g. 20,000 of 300,000 concepts). Problems:
  - Deciding which concepts to use is no small effort
  - The variation in the subset of concepts used will cause data interoperability problems
- CORE Problem List Subset of SNOMED CT
  - Represents the SNOMED CT concepts that are frequently used in problem lists
  - Derived from the UMLS-CORE data together with the dataset submitted by Beth Israel Deaconess Medical Center (most terms already in UMLS)

SNOMED CT mapping rules (1)

- Only current concepts (concept status = 0)
- Exclude Read code legacy concepts e.g. '[D] Left upper quadrant pain'
- Exclude non-human concepts
- 4 principle hierarchies used:
  - Clinical finding
  - Procedure
  - Situation with explicit context
  - Events

SNOMED CT mapping rules (2)

- Favor disorder over finding e.g. ‘Epistaxis (disorder)’ over ‘Bleeding from nose (finding)’
- Procedures included
  - ‘S/P procedure xyz’ mapped to ‘Procedure xyz’ if no appropriate concepts exist
  - Medical device mapped to procedure of inserting the device e.g. ‘Heart Valve Prosthesis’ mapped to ‘Prosthetic replacement of heart valve (procedure)’
- In a small number of cases (< 50), concepts outside the 4 principle hierarchies used e.g. ‘Thymoma (morphologic abnormality)’ and ‘Donor for liver transplant (person)’
  - Most of them were submitted to IHTSDO for addition and replaced with the new concept when added

The CORE Problem List Subset of SNOMED CT®

Data Files
(expect a new version of the subset for each new release of SNOMED CT and the UMLS Metathesaurus)

CORE Problem List Subset Version	Derived from SNOMED CT version	Derived from UMLS Metathesaurus version
SNOMEDCT_CORE_SUBSET_201008	July 2010 International Release	2010AA
SNOMEDCT_CORE_SUBSET_201005	January 2010 International Release	2010AA
SNOMEDCT_CORE_SUBSET_201002	January 2010 International Release	2009AB
SNOMEDCT_CORE_SUBSET_200911	July 2009 International Release	2009AB
SNOMEDCT_CORE_SUBSET_200908	July 2009 International Release	2009AA
SNOMEDCT_CORE_SUBSET_200907	January 2009 International Release	2009AA

Introduction

The CORE Problem List Subset of SNOMED CT® is an output of the UMLS CORE Project (CORE stands for **C**linical **O**bservations **R**ecording and

Data in the file

Column name	Example	Meaning and use
SNOMED_CID	38341003	SNOMED conceptId
SNOMED_FSN	Hypertensive disorder, systemic arterial (disorder)	SNOMED fully specified name – only included for readability, not intended for direct use in display
SNOMED_CONCEPT_STATUS	Current	SNOMED concept status
UMLS_CUI	C0020538	UMLS concept identifier – can be used to identify additional synonyms for indexing and display e.g. Hypertension, High blood pressure, Systemic hypertension, HTN, HT, HBP
OCCURRENCE	7	Number of institutions using this concept
USAGE	3.0865	Average usage volume – can be used for ranking of search results
FIRST_IN_SUBSET	200907	First appearance in subset version
IS_RETIRED_FROM_SUBSET	False	Retired from subset or not
LAST_IN_SUBSET		Last version in subset
REPLACED_BY_SNOMED_CID		Replaced by this concept in subset

Intended uses of the CORE Subset (1)

- As starter set to build problem list terminologies. Advantages:
  - Save effort of selecting the useful concepts from SNOMED CT
    - SNOMED CT contains 300,000 concepts, not every concept can be used in the problem list e.g. body structure
    - most problem list terminologies contain < 30,000 terms
  - Avoid unintentional variation due to subtle differences e.g. Epistaxis (disorder) vs. Bleeding from nose (finding)

Intended uses of the CORE Subset (2)

- Existing problem list terms can be mapped to the CORE concepts – improve data interoperability
- To extend local terminology to suit user needs
  - Add other SNOMED CT concepts
  - Post-coordination
    - can use CORE concepts as focus e.g. Kidney stone + Left => Left kidney stone
    - new concepts maintain link to CORE concepts – facilitates data aggregation
    - Can detect equivalence with new SNOMED CT concepts in the International Release

Other uses of the CORE Subset

- Priority concepts for mapping projects e.g. ICD-10, ICD-10-CM
- Quality assurance of SNOMED CT
  - QA project of IHTSDO: 'Reviewing the quality of SNOMED CT content from the perspective of fitness for purpose' – will review a sample of the CORE subset for quality of modeling, correctness and completeness of terms and relationships etc.
- NLM's MedlinePlus Connect API
  - a Web application that matches ICD-9-CM or SNOMED CT CORE Problem List Subset codes to related MedlinePlus consumer health information (ties in with 'meaningful use' criteria of EHR)

Update of the CORE Subset

- Updated with each SNOMED CT and UMLS release (4 times per year)
- Reasons for change
  - Changes in SNOMED CT and UMLS
  - Addition of new concepts e.g.
    - new dataset incorporated
    - previously missed concepts
    - new SNOMED CT concepts (e.g. some requests generated by the CORE project)
  - Retirement of concepts e.g.
    - Concept retired from SNOMED CT
    - Better concepts became available

Mapping from SNOMED CT to other terminologies or classifications

Why mapping?

- Different terminologies exist for different purposes (e.g. ICD9CM, SNOMED CT)
- 'Code once, use many times' - when data encoded for one purpose is reused for another purpose, some mapping is required
- ↑ amount of biomedical data available in electronic form =>

  ↑ demand to reuse data =>

  ↑ need for inter-terminology mapping

Some use cases for inter-terminology mapping

Primary purpose of data	Secondary use	Mapping requirement
Clinical problems list	Service reimbursement	SNOMED CT => ICD9CM
Clinical problems list	Public health reporting	SNOMED CT => ICD10
Documentation of adverse drug reactions	Reporting to regulatory institution	SNOMED CT => MedDRA
Clinical documentation	Literature search for decision support	SNOMED CT => MeSH

SNOMED CT to ICD-10 Map

- Mapping Special Interest Group and Project Group established under IHTSDO in 2008
- Co-development of a SNOMED CT to ICD-10 map by the IHTSDO and the World Health Organization (WHO) to support the epidemiological, statistical and administrative reporting needs of the IHTSDO member countries and WHO Collaborating Centers
- To support a semi-automated coding of ICD-10 classification data from a clinical record which is clinically encoded in SNOMED CT

Typical scenario of use

- Patient Jones is being discharged from the hospital. The attending physician has maintained a diagnosis and health-related problem list coded in SNOMED CT during the stay and updates the entries at discharge.
- The vendor software employs the Map, which uses a knowledge-based algorithm of sequential computable Map Rules. These rules evaluate context (data recorded about the patient in the electronic health record) and co-morbidities in the electronic record to identify the most appropriate candidate ICD-10 code list based on ICD-10 exclusion / inclusion guidance and other conventions.
- The ICD-10 coding professional later reviews and edits the classification list prior to submission for statistical morbidity reporting.

Expected benefits

- Improve efficiency of ICD-10 coding
- Better ICD-10 coding reproducibility
- Promote use of SNOMED CT
- Facilitate mapping of SNOMED to national ICD-10 extensions
- Improve quality of the two terminologies

Scope of project

- SNOMED is huge with 300,000 active concepts, not everything is relevant to ICD10 mapping
- 3 hierarchies are in scope
  - Clinical finding
  - Event
  - Situation with explicit context
- Still over 100,000 concepts in scope

The priority subset for ICD10 mapping

- For Phase One of the mapping project –
  - A subset of 9,800 SNOMED concepts has been identified
  - These concepts have high usage and will give immediate value to users
- Two sources:
  - CORE Problem List Subset - derived from real clinical usage data
  - Derived from high usage ICD10 codes

Derived from ICD10 codes

- 5,000 top ranking ICD10 codes collected from each of 5 countries: UK, Canada*, Sweden, Australia and New Zealand

  *Canadian data (ICD10CA) – only use codes mapped to the core ICD10

- Union of datasets = 6,500 ICD codes – ‘reverse mapping’ to SNOMED
  - UMLS cross-walk
  - Lexical mapping
  - Manual review

Map assumptions

- Exclusion of implied context – no assumed context or modifying semantics will be inferred beyond the definition asserted by the fully specified name and the SNOMED defining relationships, excluding qualifiers.
- Full semantic (not lexical) mapping – mapping is based on the meaning of a SNOMED concept
- Cardinality - one SNOMED source concept can map to zero-to-many ICD10 codes (seldom more than 3)

Concepts excluded from mapping

- Not all concepts in the priority set can be meaningfully and usefully mapped to ICD10
- Should be relatively few
- Main reasons for exclusion from mapping:
  - Ambiguity
    - Conflict between fully specified name and defining relationships
    - ICD10 ambiguity
  - Out of scope
    - e.g. normal findings, suspected conditions
  - Incomplete modeling

Rule-based mapping

	At run-time resolve to	Final map
◆ Map group 1		
● Rule 1 -> target code 1	code1	Codes 1 + 4 +6
● Rule 2 -> target code 2		
◆ Map group 2		
● Rule 1 -> target code 3	code4	
● Rule 2 -> target code 4		
◆ Map group 3		
● Rule 1 -> target code 5	code6	
● Rule 2 -> target code 6		

Multiple ICD codes – dagger and asterisk

- Source concepts which map to ICD chapters with asterisk conventions will be mapped to two target codes.
- The asterisk classification will always be the second map group

Source concept: 111900000 Pneumonia in aspergillosis (disorder)

Map group 1

Rule 1 -> B44.1 Other pulmonary aspergillosis

Map group 2

Rule 1 -> J17.2 Pneumonia in mycoses

Multiple ICD codes – External causes

- The source concepts denoting a condition with an identifiable cause within scope of ICD-10 chapter XX(20) will be mapped to two target codes.
- The external cause code will be assigned to the second target record.

Source concept: 242012005 Thermal burns from lightning (disorder) :

Map group 1

Rule 1 -> T30.0 Burn of unspecified region

Map group 2

Rule 1 -> X33 Victim of lightning

Map rule – gender

- Source concept: 8619003 Infertile (finding)
- Map group 1
  - Rule 1 IFA 1086007| FEMALE (FINDING) | -> N97.9 Female infertility, unspecified
  - Rule 2 IFA 248153007 | MALE (FINDING) | -> N46 Male infertility
  - Rule 3 OTHERWISE TRUE -> NULL

Map rule – age of onset

- Source concept: 32398004 Bronchitis (disorder)
- Map group 1
  - Rule 1 IF 44518008 | AGE AT ONSET OF CLINICAL FINDING (OBSERVABLE)| < 15 YEARS -> J20.9 Acute bronchitis, unspecified
  - Rule 2 OTHERWISE TRUE -> J40 Bronchitis not specified as acute or chronic

Example of map rule – acquired vs. congenital

- Source concept: 367403001 Pyloric stenosis (disorder)
- Map group 1
  - Rule 1 IFA 204671009 | CONGENITAL PYLORIC STENOSIS (DISORDER) | OR DESCENDANTS -> Q40.0 Congenital hypertrophic pyloric stenosis
  - Rule 2 OTHERWISE TRUE -> K31.1 Adult hypertrophic pyloric stenosis

Map advice

- Carries additional textual guidance to clarify the map's rationale and usage
  - ADDITIONAL CODES FROM *XX, YY-ZZ NOT MAPPED*
  - FOURTH CHARACTER REQUIRED TO IDENTIFY PLACE OF OCCURRENCE
  - THIS IS AN EXTERNAL CAUSE CODE FOR USE IN A SECONDARY POSITION
  - POSSIBLE REQUIREMENT FOR CAUSATIVE DISEASE CODE
  - POSSIBLE REQUIREMENT FOR AN EXTERNAL CAUSE CODE
  - POSSIBLE REQUIREMENT FOR MORPHOLOGY CODE
  - MAPPED WITH WHO GUIDANCE
  - MAPPED WITH IHTSDO GUIDANCE

Mapping heuristics and exemplars

- To ensure the resulting maps are reproducible:
  - Mapping heuristics - a set of guidelines for the choice of target codes and creation of map rules
  - Map exemplars – for documentation and training
  - 3 rounds of mapping prototype exercise to refine and improve guidelines

Quality assurance

- Three-tiered quality assurance
  - Independent map creation
    - Each concept mapped by 2 map specialists
    - If the maps are different -> conflict resolution procedure
  - Review of sample of final maps by independent party
  - Independent study of use of map
    - Based on a collection of anonymized clinical records, ICD-10 codes will be generated separately by conventional manual coding and by semi-automated coding making use of the Map
    - The 2 methods compared for efficiency, accuracy and variability

SNOMED CT to ICD-9-CM Map

- ICD-9-CM codes required for reimbursement in the U.S.
- General equivalence cross-maps
  - Created and maintained by IHTSDO
  - Updated and released with the International Release
  - General 1-1 or 1-N maps
- Rule-based reimbursement map
  - Commissioned by NLM
  - Rule-based map to support reimbursement coding
  - Draft of 5,000 maps available

SNOMED CT to ICD-10-CM Map

- ICD-10-CM codes will be required from October 2013 onwards
- NLM is planning to create a map to ICD-10-CM
  - Will be drawing on experience and tools developed for the international ICD-10 mapping effort
  - Added complexity because of finer granularity of ICD-10-CM (70,000 vs. 10,000 codes)
  - First batch of SNOMED CT concepts – CORE Subset

Other maps

- MEDCIN to SNOMED CT
  - MEDCIN is a proprietary clinical terminology used by the U.S. Department of Defense and some EHR vendors
  - Map available in the UMLS
- SNOMED CT to MedDRA
  - MedDRA (Medical Dictionary for Regulatory Activities) is for pharmaceutical companies to report adverse effects of medications
  - Used by U.S. Food and Drug Administration and the European Union
  - Map discussed but no definite plan yet