

Managing clinical terminology in the post-classification era

James R. Campbell

University of Nebraska Medical Center

June 1, 2008



Or...

Extending the T-box in Clinical Real Time

James R. Campbell

University of Nebraska Medical Center

June 1, 2008



Disclosure

My employer, the University of Nebraska, receives financial support from General Electric for provision of the vocabulary services and utilities I will describe during this talk



Learning Objectives

- Identify the important features and limitations of classifications for implementation of meaning in clinical information systems
- Understand the advantages of an ontology relative to a classification for purposes of:
 - clinical expression,
 - permanence
 - decision support
 - interoperation
- Appreciate the features of SNOMED CT which provide functionality for 21st century clinical information systems
- Enumerate the challenges facing deployment of SNOMED CT when replacing clinical classifications



Definitions

- **Classification:** a controlled vocabulary, generally employing taxonomic features, which specifies assignment for every element of a domain to a single category within the classification
- **Reference terminology:** A comprehensive, consistent and logically organized set of concepts and associated terms (descriptions), designed to completely embody the expressive detail of a given domain, supported by a set of relationships which define the elements within the domain, and explain their meanings relative to each other
- **Ontology:** A reference terminology which specifies a formal definition of meaning (a conceptualization) for one or more related domains of discourse and provides a comprehensive set of computable definitions for concepts within those domains



Reference Terminology vs Ontology

- Reference terminologies are concept-based coding systems which employ principles of ontologies, however...
- Pragmatic choices for health care information systems may require descriptive or knowledge-based elements that are not properly definitional
- Decision support requirements, especially subsumption relationships for clinical use cases, may often be epistemologic rather than ontologic in nature
- Furthermore, requirements for expression of uncertainty and extended context must be met within the constraints of the vendor information model



Controlled Vocabularies

- Classifications:
 - ICD-10XX; 1978:ICD-9-CM; CPT; HCPCS
- Reference terminologies:
 - 1994: MED; 1995: LOINC; 1999: SNOMED RT; NDF-RT; 2002:SNOMED CT
- Ontologies:
 - 1991:Galen, FMA, Gene Ontology



Requirements for Successful Implementation of Clinical Terminology

- Expressive:
 - Comprehensive content
 - Unambiguous
 - Provides variable semantic granularity
 - Supports parochial expression and terms
- Permanence
- Data re-use:
 - Clinical data record must support information needs of multiple users
- Decision analytical:
 - Supports aggregation of clinical reference groups
 - Includes relevant declarative domain knowledge
- Interoperable:
 - Reliably supports machine-based transmission and integration of conceptual content
- FREE OF CHARGE



Clinical vocabulary history: studies in classifications

1980's> Historical efforts at implementation of clinical information systems attempted to align the reimbursement, statistical reporting and clinical descriptive requirements by employing classifications for all clinical coding



Experience with classifications

- Expressivity:
 - ICD-9-CM and CPT covered <65% of portions of clinical record content
- Permanence:
 - Elusive meaning of “NEC”
- Data re-use:
 - Granularity appropriate for reimbursement/epidemiology did not support translation or mapping
- Decision analytical:
 - Record retrieval for ICD-9-CM 58%
- Interoperability:
 - Supports token interoperability (concept enumeration) supported by a synchronous update process with annual publication cycle

Elkin PL, Ruggieri AP, Brown SH, Buntrock J, Bauer BA, Wahner-Roedler D, Litin SC, Beinborn J, Bailey KR, Bergstrom L. A randomized controlled trial of the accuracy of clinical record retrieval using SNOMED-RT as compared with ICD9-CM. Proc AMIA Symp. 2001:159-63.

Clinical Interoperability

- 1) **Functional**: physical link establishes communication
- 2) **Transactional**: sharing message data
- 3) **Semantic**: sharing data meaning

Semantic interoperability refers to the ability of computer systems to exchange data in such a way that the defining features of the data are software accessible and the system can employ the information meaningfully

- 4) **Procedural**: sharing meaning within a shared executable process
- 5) **Ergonomic**: software employed within a shared work plan



Limitations of clinical classifications

- No finite list of classification terms can be prepared which will accurately and comprehensively capture clinical recording needs across all realms
- Synchronous publication of classification data is not responsive to local and changing recording needs
- Local variation in expression is not supported with authoritative publication
- Fixed granularity of statements is not clinically acceptable
- Permanence is not supported with most designs
- Decision support is limited



Clinical Vocabulary: Conclusions

- Compositional formalisms with well defined models of meaning are required
- Algorithms for creation of extensions must be supported, understandable and reproducible for interoperability
- Decentralized authoring and communication of content are needed for full expression responsive to local requirements



Operational features of an ontology

- Conceptual model of meaning supports local (and ubiquitous) authoring
- Definitional semantic interoperation with detailed machine interpretable meaning is possible
- Granularity of expression need not be controlled
- Asynchronous sharing of content is possible
- Release cycles need not be restricted



SNOMED CT operational features

- Studies of content document best clinical comprehension and expression
 - Includes vestiges of legacy systems (including classifications) which are slowly being resolved
- Model of meaning in development
- Formalisms for testing of equivalence and subsumption have been developed but not operationally tested



Semantic roots...SNOMED CT

- Clinical finding
- Body structure
- Procedure
- Substance
- Organisms

- Finding site
- Associated morphology
- Associated with
- Severity
- Interprets
- Has interpretation

“Hair tourniquet (Clinical finding)”|15420001000004107 | [Defined]
 { 116680003 | is a | = 233958001|peripheral ischemia |,
 363698007 | finding site | = 120573002 | extremity part |,
 116676008| assoc morphology| = 71173004 | compression|,
 246705003 | causative agent | = 312025007 | human hair }

- Pharmaceutical product

- Clinical course.....



Semantic roots...SNOMED CT

- Clinical finding
- Body structure

- Procedure site
- Procedure morphology
- Method

- **“ORIF right ankle fracture(Procedure)”|15810001000004109 | [Defined]**
116680003 | is a | = 71388002| procedure |,
- **Group 1**
{260686004 | method | = 426530000 | fixation |
(405813007 | Procedure site – direct | = 33696004 | bone structure of ankle|:
272741003 | laterality | = 24028007 | right |),
363700003 | direct morph | = 72704001 | fracture |,
424226004 | using device | = 31031000 | orthopedic internal fixation system }
- **Group 2**
{260686004 | method | = 426530000 | open reduction |
(405813007 | Procedure site – direct | = 33696004 | bone structure of ankle|:
272741003 | laterality | = 24028007 | right |),
363700003 | direct morph | = 72704001 | fracture|}

- Using energy.....

Semantic roots...SNOMED CT

- Associated with
- Causative agent
- Due to
- After
- Occurrence.....

- Physical force
- Specimen
- Observable entities
- Events
- Environments and geographical

- Organisms
- Physical objects

- **“Cadmium exposure (Event)”|12840001000004100 | [Defined] scales**
116680003 | is a | = 420239005 | exposure to toxin
{246705003 | causative agent | = 66586000 | cadmium}



Reference Terminologies

Barriers: in Clinical Systems

- Costs of development
- Costs of implementation
- Uncertainty among software vendors:
 - Hard to understand
 - Unclear: how to utilize?
 - Permanency, how to manage?
 - No examples of best implementation practice



Ten years of experience with SNOMED CT

- 1997: Implementation of problem list with SNOMED International V3
 - No formal model of meaning
 - Confusion regarding management of context
 - Comprehension ~ 80%
- 1999: SNOMED RT
 - First formalisms for model of meaning
- 2002: SNOMED CT
 - Comprehension ~ 92%
 - Formalisms for interoperable use of semantic definitional model



Nebraska Lexicon© for GE

- Implementation principles:
 - Fully capture and encode content of any clinical utterance requested by clinicians employing SNOMED formalisms; use pre-coordinated content when available, otherwise post-coordinate as required
 - Employ pragmatic set of SNOMED terms for lookup and expand entrance (interface) terms as required by the clinician for ease of clinical navigation and selection
 - Maintain current map from each concept to ICD-9-CM for billing and epidemiological reporting
- Post-coordination at the user interface has been managed using “Just-in-time” administrative protocols



“Just-in-time post-coordination”

- If clinician cannot identify the required clinical phrase when making an entry into the problem list; records “Unlisted problem” in the record with a full text description
- Clinical coding team obtains weekly report of entries
- Clinical utterance is evaluated:
 - Disambiguated
 - Compared to SNOMED content
 - Mapped to ICD-9-CM
- New terms are added to the interface and pointed to the SNOMED CT concept identifier in the dictionary tables
- New concepts are modeled as extensions to SNOMED CT in an extension namespace which is maintained with a description logic classifier
- The “Unlisted problem” pointer in the database is updated to a dictionary reference with the complete and accurate content
- The problem list history files maintain a record of the original entry in addition to the fully modeled and coded content now in the record



Semantic Roots of Problem List

- Clinical findings (disorders and findings)
- Situations with explicit context
- Events
- Procedures (history)



Nebraska Lexicon© for GE

- UNMC has been employing SNOMED standards since 1997 and supporting problem list terminology for GE enterprise customers for seven years
- 23 healthcare institutions nationwide use the Lexicon interface subset of SNOMED CT for their problem list implementation
- UNMC delivers term sets, ICD mapping and post-coordinated content to support interoperability needs twice annually or as needed by the enterprise
- 2007 problem list SNOMED subset 10,741 concepts with 8.6% post-coordination; 16,545 entrance terms



Evaluating expressiveness

- Within a model of meaning, a “degree of freedom” represents a knowledge token that must be created and maintained. It is a fundamental unit of effort in creation and maintenance of an ontology or reference terminology.
- For SNOMED CT:
 - The semantic degrees of freedom of an (extension) concept is the sum number of the defining and qualifying relationships that are employed in the long normal form of the concept definition
 - The lexical degrees of freedom of a concept is the sum number of terms that are employed in the clinical interface which reference the concept



Counting Degrees of Freedom

1 Lexical degree of freedom

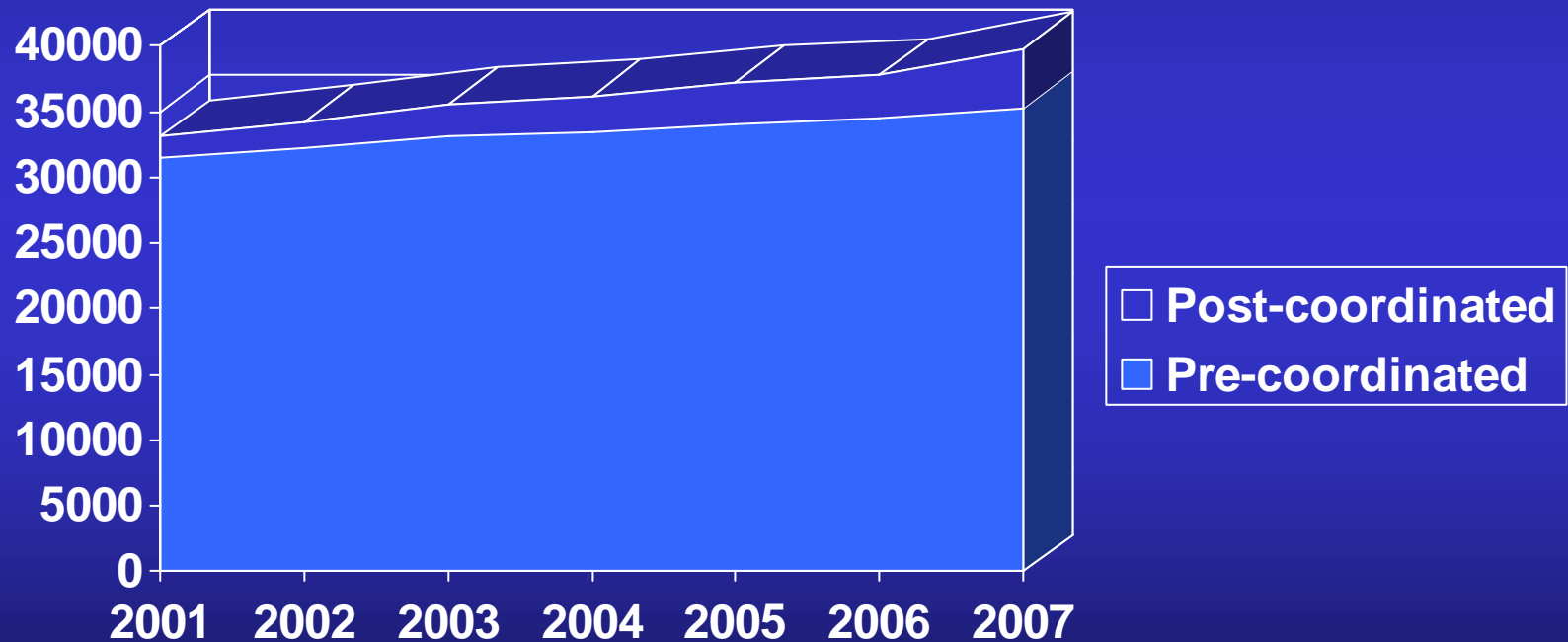
“ORIF right ankle fracture(Procedure)”|15810001000004109 | [Defined]

- 1 { 116680003 | is a | = 71388002 | procedure |,
- 2 Group 1
- 3 260686004 | method | = 426530000 | fixation |
- 4 405813007 | Procedure site – direct | = 33696004 | bone structure of ankle|,
- 5 272741003 | laterality | = 24028007 | right |,
- 6 363700003 | direct morph | = 72704001 | fracture |,
- 7 424226004 | using device | = 31031000 | orthopedic int fixation system
- 8 Group 2
- 9 260686004 | method | = 426530000 | open reduction |
- 10 405813007 | Procedure site – direct | = 33696004 | bone structure of ankle|,
- 11 272741003 | laterality | = 24028007 | right |,
- 12 363700003 | direct morph | = 72704001 | fracture } }

Semantic degrees of freedom



Lexicon: Semantic DOF



2007 statistics: 9822 concepts from SNOMED pre-coordination
919 Nebraska concepts in post-coordination
438 Nebraska post-coordinated concepts retired
10741 concepts in the subset



Failures of Clinical Expression

- Majority of post-coordination issues for problem list are required for statements of clinical specialization:
 - “Increased wound drainage”
 - “History of fecal impaction”

Majority of primitive Lexicon concepts relate to infrequent deficiencies in SNOMED model of meaning:

“HOMOZYGOUS C6777T METHYLENETETRAHYDROFOLATE REDUCTASE DEFICIENCY (DISORDER)”

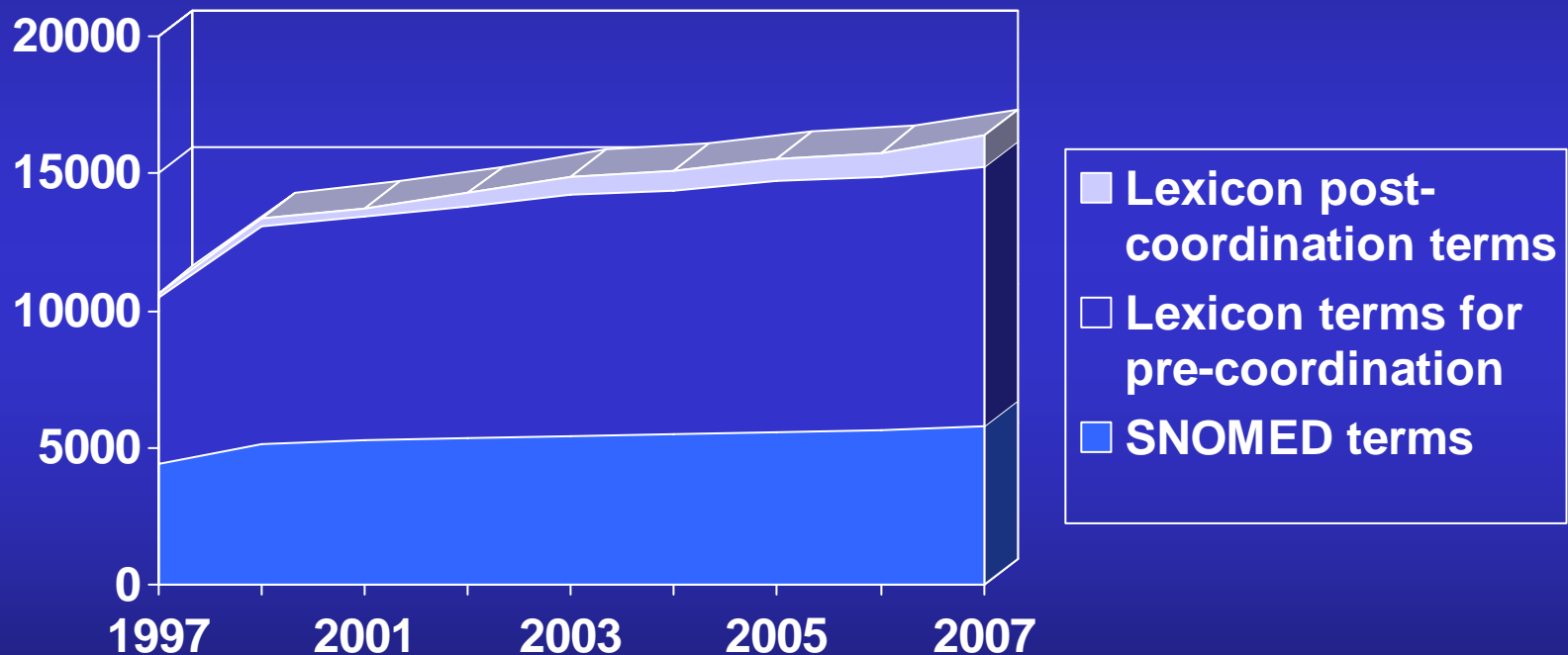
“PATIENT HAS BEDTIME ROUTINE (FINDING)”

“HOME NASAL BIPAP USE (PROCEDURE)”

- Lexicon contains only 9% primitives



Lexicon: Lexical DOF



**2007 statistics: 5778 terms from SNOMED pre-coordination
9452 Nebraska terms for pre-coordinated SNOMED
1152 terms for Nebraska post-coordination
16545 terms in the interface set**



Lexicon© Report Card

- Expressivity:
 - Primary reason for continued success of Lexicon implementation is timely delivery of meaningful terms within consortium of shared use
 - Authoring of post-coordinated content is a secondary and tractable issue
 - Post-coordination by clinician faces problems except in restricted cases
- Data re-use:
 - Complete and current map to ICD-9-CM has been an ongoing requirement and important to implementation at majority of sites
 - Contention between clinical and reimbursement use cases is largest ongoing editorial problem
- Permanence:
 - Managed within SNOMED CT history mechanisms and extension management utilities
- Decision analytical:
 - Preliminary experiments employing SNOMED extension integrated with clinical record have demonstrated superior utility of deployment
- Interoperability:
 - Challenges are substantial and relate
- Costs:
 - Acceptable and well managed by resources within community of shared use



Permanence: SNOMED CT History Management

- Each SNOMED release cycle exposes editorial revisions:
 - New pre-coordinated content must be compared against extension namespace content
 - Editorial changes participating in definitions of post-coordination must be reviewed
- Clinical definition of any problem list entry must never be allowed to change, although the SNOMED concept identifier may do so
- Average annual retirement of post-coordination 6%

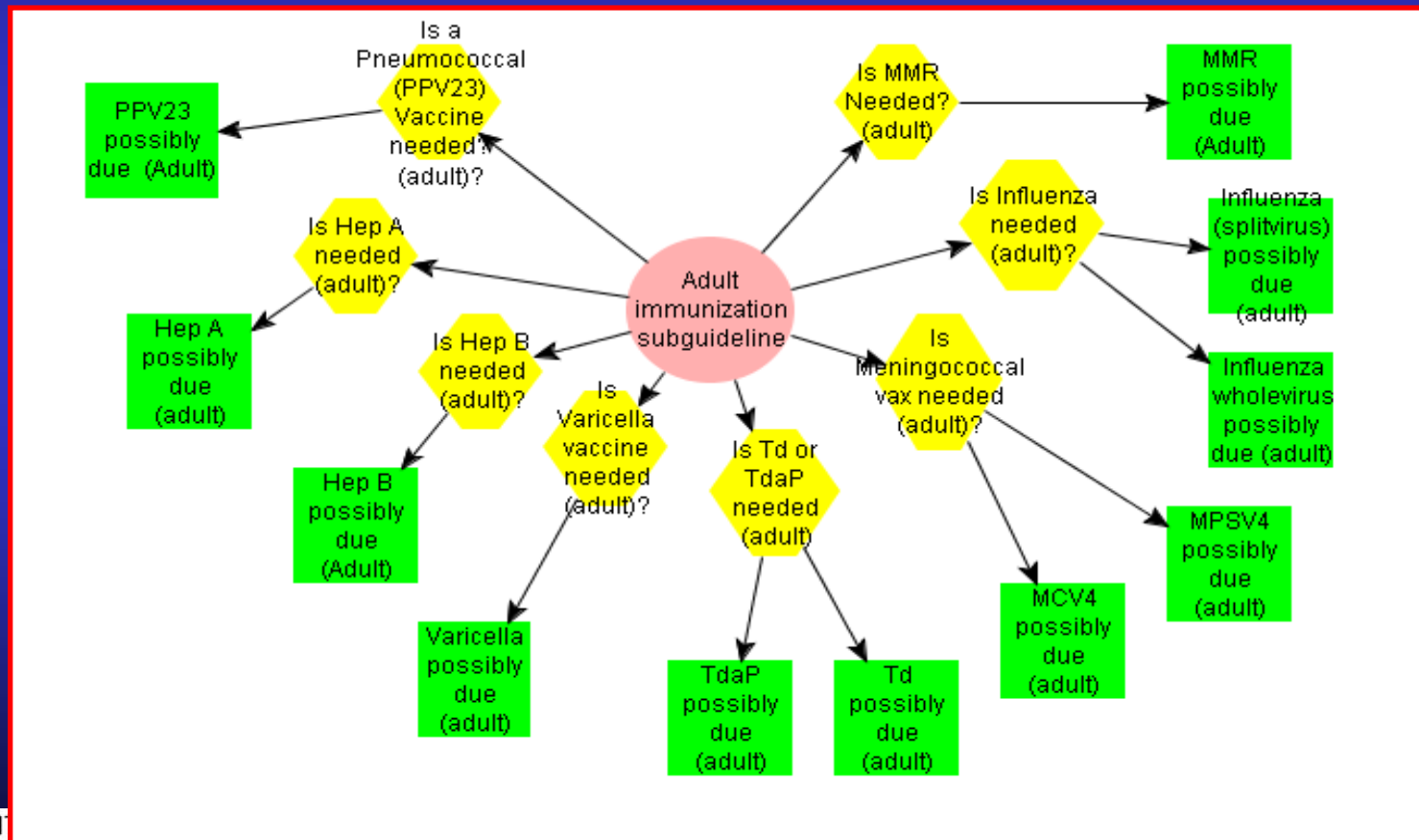


Evaluating Decision Support

- As part of the SAGE guideline engine development, we evaluated utility of the Lexicon problem list for decision support
- Frame based knowledge modeling of CDC immunization guidelines was accomplished with Protégé
- An API linking the decision engine to the GE information model queried the EHR database including Lexicon problem list
- Criteria were modeled within decision models which reproduced the source guideline logic



Characteristics of the Immunization Knowledge Model



Characteristics of the Immunization Model

- Entire text of CDC Immunization Guideline distilled into 75 individual “IF-THEN” structured logic statements supporting three clinical scenarios
 - Vaccine advice at birth
 - Primary care office visit
 - Population based reminders
- 197 conceptual references were identified as required for decision making from the source guideline

<http://sage.wherever.org>



Pre-coordinated Concepts by Semantic Class

SNOMED domain	n
Situation with explicit context	3
Disorder	51
Finding	23
Observable entity	10
Occupation	4
Organism	1
Person	1
Procedure	9
Product (clinical drug)	50
Qualifier	16
Racial group	1
Substance	5
Total	n = 174



Concept Inventory by vMR Query Class

vMR query	SNOMED	LOINC
Adverse reaction	42	
Agent	0	
Alert	0	
Appointment	0	
Encounter	0	
Goal	0	
Observation	39	5
Order	6	
Medication Order	24	
Problem	78	
Procedure	4	
Referral	0	
Substance administration	32	



Query Concept Inventory by Semantic Complexity

Category	n
Category 1 (Concept entity)	35 (17.8%)
Category 2 (Subsumption)	139 (70.5%)
Category 3 (Boolean constructions)	12 (6.1%)
Category 4 (Post coordination)	11 (5.6%)
Total	n = 197



SNOMED CT:

Experiments in Decision Support

- Binding a guideline decision engine to SNOMED CT domain ontology is feasible
- Expression of guideline concepts required modest vocabulary extension development
- SNOMED CT relationships offered flexible and useful support for guideline inference



Semantic Interoperation

- Definitional semantic interoperation is supported for SNOMED CT problem list by classification of extension concepts and publication of the Normal Form
- We analyzed the validity of algorithms for testing of semantic equivalence by comparing SNOMED problem lists incorporating post-coordination developed at two separate institutions

SNOMED CLINICAL TERMS: Transforming Expressions to Normal Forms. International Health Technology Standards Development Organization; Copenhagen, 2008.



Methods: Study sites

- **University of Nebraska Medical Center; Omaha, NE, USA**
- **Hospital Italiano; Buenos Aires, Argentina**

Hospital Italiano is a non-profit university hospital, associated with a large network of ambulatory care services, located in Buenos Aires, Argentina. Clinical data entry uses a local interface terminology for problem list. This interface was created from 2 million terms of free text recorded by local users in legacy systems. Every concept from the terminology has been coded with SNOMED CT. Current Hospital Italiano terminology has 24,800 concepts and more than 100,000 interface terms, with approximately 70% post-coordinated content.



Methods

- SNOMED CT problem lists were merged and compared in order to develop a shared problem list vocabulary for primary care
- The merged concept inventory was processed with a classifier compliant with SNOMED guidance in order to compute Long Normal Forms for each concept



Most Common Problems

Nebraska

- Health maintenance
- Hypertension
- Upper respiratory infection
- Otitis media
- Sinusitis
- Allergic rhinitis
-

Argentina

- Maintenance of health
- Arterial hypertension
- Nicotine dependence syndrome
- Dyslipidemia
- Fever
- Infection of upper airways

“Upper respiratory infection (Clinical finding)”|54150009 | [Defined] 116680003 | is a | = 64572001| disease |

{363698007 | finding site | = 58675001 | upper respiratory tract structure | ,246705003 | causative agent | = 36272005 | infectious agent }



Methods

A convenience sample of the classifier output was systematically evaluated by the two authors:

- 1) verifying the intended meaning in the original concept term
- 2) verifying SNOMED editorial guidance for modeling of the meaning
- 3) validating the post-coordinated concept against published SNOMED editorial principles
- 4) verifying true synonymy of content where asserted by the classifier
- 5) assessing for “hidden synonymy” against concepts classified as unique; requiring intensive review of related concepts and alternative post-coordination strategies in the merged set



Results

	Nebraska	H Italiano	Merged
Pre-coordinated	9734 (94.3%)	7666 (23.7%)	14069 (35.7%)
Post-coordinated	585 (5.7%)	24727 (76.3%)	25312 (64.3%)
Total concepts	10319	32393	39381

Problem list subsets at the two institutions were notably different in concept inventory and frequency of post-coordination



Results: Problem List Semantics

	Nebraska	H Italiano	Merged
Disorders	7467 (72%)	15891 (65%)	20633 (66%)
Findings	1372 (13%)	3516 (27%)	7328 (23%)
Events	30 (0.3%)	82 (0.3%)	104 (0.3%)
Procedures	1098 (10.6%)	112 (0.5%)	1201 (3.8%)
Situation w/ explicit context	339 (3.2%)	864 (3.5%)	1178 (3.7%)
All other	(0.9%)	(3.7%)	(3.2%)

Semantics displayed different patterns of recording; there was only 20% overlap in the pre-coordinated concept sets



Results

Failure of post-coordination	
Vague source concept utterance	3 (1.2%)
Limited expressiveness of SNOMED model	3 (1.2%)
Conflict of SNOMED guidance	9 (3.6%)
Non-compliance with SNOMED editorial guidelines	72 (28.85)
Subtotal of classification errors	87 (34.8%)
Masked synonymy	3 (1.2%)

N=250 concepts



Semantic Interoperation

- Clinically driven problem list recording at these two institutions led to development of subsets with substantially different lexical and semantic features
- SNOMED Normal Forms were useful for comparing disparate data sets, but would fail 1/3 of the time in this trial
- The greatest deficiency we identified in post-coordination was inconsistent behavior in the concept modeling at the two institutions, largely due to lack of education and training



Costs:

Post-coordination maintenance

- Resolving ambiguity 1.0 FTE \$40,000
- Checking record
- Coding ICD-9-CM
- Maintain tables
- Model post-coordination 0.3 FTE \$25,000
- Classifier updates \$20,000

Entirely covered by contract arrangements
within user consortium



Lexicon© Report Card

- Expressivity:
 - Primary reason for continued success of Lexicon implementation is timely delivery of meaningful terms within consortium of shared use
 - Authoring of post-coordinated content is a secondary and tractable issue
- Permanence:
 - Managed within SNOMED CT history mechanisms and extension management utilities
- Data re-use:
 - Complete and current map to ICD-9-CM has been an ongoing requirement and important to implementation at majority of sites
- Decision analytical:
 - Preliminary experiments employing SNOMED extension integrated with clinical record have demonstrated superior utility of deployment
- Interoperability:
 - Challenges are substantial and relate heavily to formalization, education and training in model of use
- Costs:
 - Acceptable and well managed by resources within community of shared use



Operational challenges employing SNOMED CT in EHR

- Variation in local expression and institutional requirements demand scalable procedures for post-coordination and extension term management
- Differences must be resolved with allied terminologies in support of extended model of meaning across Observables and Products
- Maps to classifications and other references must be available and useful



Requirements for SNOMED clinical success

- Education of vendor community
- Tooling for extension set maintenance:
 - Post-coordination templates enforcing model of meaning
 - Inexpensive full function classifier
 - Modular tools for release management of large extensions
- (Emergence of middle-ware brokers to manage services for communities of users)
- Creation of validated “starter subsets” for clinical recording use cases
- Completion and confirmation of SNOMED model of meaning
- Education and training in employment of post-coordination
- Development of vendor bindings to SNOMED CT as domain ontology
- Development of high quality maps to meet the varying use cases across realms
- Emergence of interoperable knowledge-based systems employing SNOMED CT as domain ontology



Conclusions

- SNOMED CT is clinically expressive
- SNOMED CT requires local interface management
- SNOMED CT is not complete
- Paradigms for decentralized, collaborative, responsive and interoperable vocabulary management must be developed
- Vendors and users are unfamiliar with and have fears about composition and post-coordination
- Study, education and communication are required



Discussion?

