SNOMED CT® as the clinical terminology for representation of the clinical care of the newborn infant

Andrew G. James MBChB FRACP FRCPC

A Capstone Project

Presented to the Department of Medical Informatics and Clinical Epidemiology, School of Medicine, Oregon Health & Science University

in partial fulfillment of the requirements for the degree of

Master of Biomedical Informatics

May 2007

School of Medicine

Oregon Health & Science University

Certificate of Approval

This is to certify that the Master's Capstone Project of

Andrew G. James

**SNOMED CT® as the clinical terminology for representation
of the clinical care of the newborn infant**

Has been approved

_____

Kent A. Spackman MD PhD, Capstone Advisor

Table of Contents

List of Tables

Acknowledgements

Abstract

Background: The Systematized Nomenclature of Medicine Clinical Terms®
(SNOMED CT) is the most sophisticated reference terminology currently available for the
representation of health care. In several countries it is being adopted as the terminology
for the documentation of clinical care.

SNOMED CT has evolved from the Systematized Nomenclature of Pathology [SNOP,
1965] into the current Systematized Nomenclature of Medicine [SNOMED, 2002] over
forty years. The content of the terminology has been determined largely by the voluntary
contributions of many, diverse clinical groups. An unforeseen consequence of this
evolutionary process may be that some unique, clinically relevant concepts of highly
specialized clinical domains are not represented as SNOMED CT concepts.

The purpose of this project was to assess the performance of SNOMED CT as the
terminology for the representation of the clinical care of the newborn infant.

Methods: The SNOMED CT July 2006 release was systematically examined using the
CliniClue 2006 Terminology Browser available for this release to determine the degree of
representation for 881 terms that are relevant for the clinical care of the newborn infant by
SNOMED CT concepts.

The 881 terms were extracted from the Clinical Information Management System (CIMS)
used in the Neonatal Intensive Care Unit at the Hospital for Sick Children, Toronto,
Ontario, Canada. The CIMS terms were categorized as an observation, diagnosis,
procedure or drug. The representation for each CIMS term by a SNOMED CT concept
was assessed as absent, partial or complete representation by a single clinician.

Results: Ninety-six percent of the 881 CIMS terms that describe an observation, diagnosis, procedure or drug are represented within the SNOMED CT July 2006 release. There is complete representation for 86.4% of the terms, partial representation for 10.2% of the terms and no representation for 3.4% of the terms.

Table 1. Representation of CIMS terms by SNOMED CT concepts

Category of CIMS term	Number of CIMS terms	No representation	Partial representation	Complete representation
Diagnosis	434	2.8%	6.4%	90.8%
Observation	172	4.1%	6.4%	89.5%
Procedure	110	4.6%	14.6%	80.9%
Drug	165	3.6%	21.2%	75.2%
All terms	881	3.4%	10.2%	86.4%

Ninety-seven percent of the 761 CIMS terms that are completely represented by a SNOMED CT concept are represented by a single, pre-coordinated SNOMED CT expression. Ninety-three percent of the SNOMED CT concepts that completely represent CIMS terms have a current Concept Status. Twenty-eight percent of the SNOMED CT concepts are fully defined SNOMED CT concepts and 49.4% of the SNOMED CT concepts have one or more synonyms.

Conclusion: SNOMED CT is a comprehensive reference terminology that provides robust, structured representation based on description logic for the majority of the CIMS terms that are relevant for the clinical care of the newborn infant.

Introduction

The Systematized Nomenclature of Medicine Clinical Terms® (SNOMED CT) is a comprehensive, concept-based, clinical terminology with a semantic model based on description logic that uniquely identifies and describes clinically relevant concepts.[1] SNOMED CT has a polyhierarchical structure with multiple parent-child relationships together with horizontal relationships between concepts in different subtype hierarchies that define the meaning of a SNOMED concept relative to other SNOMED concepts. SNOMED CT contains more than 370,000 active concepts and more than one million active descriptions.[2] There are nearly 1.5 million relationships in SNOMED CT. The coded information for the terminology consists of SNOMED CT expressions that are references to SNOMED CT concepts. Clinical concepts may be represented as a single, pre-coordinated SNOMED CT expression or a post-coordinated SNOMED CT compositional expression comprised of two or more SNOMED CT concepts. The use of post-coordinated compositional expressions for the representation of clinically relevant concepts increases both the comprehensiveness and flexibility of SNOMED CT as a clinical terminology. This paper describes the terminology as of the July 2006 release.

The SNOMED CT terminology model

The SNOMED concept, description and relationship are the three core components of SNOMED CT terminology model. Additional features include subsets, cross-maps, extensions and the history mechanism.[1]

SNOMED CT Concept and Description

The SNOMED concept has a single, specific and permanent meaning that represents a

1

clinically relevant idea. Each concept has two or more textual descriptions that name the concept in natural language. A SNOMED concept is defined by a set of descriptions and a set of relationships to other SNOMED concepts. The computational definition for the SNOMED concept is based upon description logic. The defining relationships may be vertical parent-child relationships that confer inheritance of characteristic features of the parent concept or horizontal relationships defined by attributes that link a SNOMED concept with one or more other SNOMED concepts.

SNOMED concepts are organized into top-level subtype hierarchies that are defined by their parent concepts and attributes.[1] The eighteen top-level hierarchies of SNOMED CT July 2006 release together with a description of concept representation for each hierarchy are appended as Appendix 1. The Clinical Finding, Pharmaceutical/Biologic Product and Procedure subtype hierarchies contain SNOMED concepts that are most frequently used for the representation of clinical care.

SNOMED concept definitions are modeled through the use of approximately 50 approved attributes that can be applied to one or more hierarchies. The approved attributes that are used in SNOMED CT July 2006 release are appended as Appendix 2.

Each SNOMED concept has a Fully Specified Name that is a unique, unambiguous, name for the concept. Each Fully Specified Name also contains a "hierarchy tag" that indicates where the concept is integrated into the primary SNOMED hierarchy. Each SNOMED concept also has a Preferred Term that is a common phrase or word used by clinicians to name that concept and may have one or more synonyms.

Each SNOMED concept has a unique semantically meaningless numeric identifier

[SNOMED Conceptid] that supports the terminology's computational capabilities. The status of a SNOMED concept may be current, retired, duplicate, outdated, ambiguous, erroneous or limited. Some concepts with the limited status have been incorporated into SNOMED CT from ICD-9-CM, ICD-10 and other terminologies.

A SNOMED concept is described as "fully defined" if its definition is sufficient to distinguish it from all of its supertype concepts. A concept is "primitive" if it is not fully defined. The existence of fully defined concepts is critical for the processes of auto-classification and the computation of equivalence and subsumption between different methods of expressing the same meaning.

SNOMED CT Relationship

Logical relationships between two or more concepts are used to define a SNOMED concept in terms of other SNOMED concepts. Every SNOMED concept is a refined type, or sub-type, of one or more other concepts. These hierarchical defining relationships are structured with description logic and have relationship type "is_a". The difference between two concepts may also be represented by other defining relationships that confirm and enhance the accuracy of the subtype hierarchy, enable concepts to be refined by increasing the specificity of a defined relationship and allow recognition of equivalence between different ways of expressing the same concept. Defining relationships may be aggregated into Relationship Groups, or role groups, to indicate the way they relate to each other. Relationship Groups clarify meaning and are used to avoid misinterpretation.

Subsets

A SNOMED CT subset is a set of references to SNOMED components that are used for a predefined, specific purpose. Subsets are used to identify SNOMED components for language and dialect support, content filtering and alternative hierarchies. The components of the subset are appropriate to a particular user, context, language, dialect, country, specialty or discipline, or organization.

Cross Maps

SNOMED CT Cross Maps show the correspondence between SNOMED concepts and those of other terminologies, classifications and coding systems. Cross Map sets are available within SNOMED CT July 2006 release for ICD-9-CM, ICD-10 (UK edition), ICD-O3, OPCS-4 (UK edition), NIC, NANDA, PNDS, Clinical Care Classification and the Omaha System.[2]

The cross mapping mechanism enables automatic mapping from one SNOMED CT concept to a single appropriate, matching code, or a single collection of codes that together represent the same concept, in another coding system. Cross mapping is rarely one-to-one and may be many-to-many. The mapping may be from SNOMED CT to another coding system or from another coding system to SNOMED CT. Cross mapping ensures that SNOMED CT can be used effectively to reference other terminologies and classifications.

Extensions

The SNOMED CT extension mechanism enables authorized users [organizations] to add Concepts, Descriptions, Relationships or Subsets to the core content of SNOMED CT to

support specialized or unique needs. The mechanism provides a defined structure to maintain the unique identification of Extensions across organizations and facilitates the submission, inclusion, use and migration of terminology developed as part of an extension into the core content of SNOMED CT. Extensions within the SNOMED CT July 2006 release include US proprietary drugs, UK proprietary drugs and UK administrative concepts.

History mechanism

The bi-annual releases of SNOMED CT facilitate graceful evolution of the terminology to support the inclusion of new developments in healthcare and error correction. The types of changes include new Concepts, Descriptions, Relationships [between Concepts], Cross Maps and Subsets together with updates and retirements of any of these components. The history mechanism documents all changes to SNOMED CT.

The evolution of SNOMED CT

SNOMED CT has evolved from the Systematized Nomenclature of Pathology [SNOP, 1965] into the current Systematized Nomenclature of Medicine Clinical Terms [SNOMED CT, 2002] over forty years. There has been a remarkable evolution from the focus on anatomical morphology of SNOP to the comprehensive representation of healthcare by SNOMED CT concepts.

SNOMED, the first version of a family of terminologies, was released in 1974 and the terminology contained concepts for etiology and location in addition to SNOP's concepts for anatomical morphology. The next three members of the SNOMED family of terminologies [SNOMED II, 1979; SNOMED version 3.0, 1993; SNOMED version 3.5,

1998] delivered an expansion of SNOMED concepts to include disorders, living organisms and social context. Veterinary content was re-integrated into SNOMED version 3.0 in 1993 and the Logical Observation Identifier Names and Codes were integrated into SNOMED in 1997.

The transformation of SNOMED version 3.5 into the SNOMED Reference Terminology [SNOMED RT, 2001] was characterized by the introduction of description logic definitions for SNOMED concepts and an auto-computed, code independent subtype hierarchy. The evolution of SNOMED RT into SNOMED CT was the outcome of the merger of SNOMED RT with the NHS Clinical Terms, the computer-based nomenclature and classification [medical terminology] of the National Health Service of the United Kingdom.

NHS Clinical Terms evolved from mnemonic, non-hierarchical codes developed by James Read for use in his general practice in the United Kingdom. The first version of the Read Code, published in 1984, was a hand crafted, code dependent, subtype hierarchy that contained approximately 10,000 terms. The Royal College of General Practitioners and the British Medical Association recommended the Read Codes as the standard terminology for general practice in 1988. The United Kingdom's National Health Service purchased the Read Codes in 1990. There was further development and enlargement of the Read Codes to capture medical concepts in a wide variety of clinical situations. Version 3 of the Read Codes (CTV3) was a controlled medical vocabulary produced during the Clinical Terms Project with the involvement of over 2,000 health care professionals from all United Kingdom specialties. CTV3 expanded from 150,000 concepts in 1992 to 250,000 concepts in 1999.

Table 2. The evolutionary history of SNOMED CT

Year	Terminology	Comments
1965	SNOP	Focus on morphology [10, 990 concepts]
1974	SNOMED	Focus on morphology, etiology and location
1979	SNOMED II	[30,547 concepts]
1984	Read Codes, version 1	Handcrafted, code dependent, subtype hierarchy [10, 000 terms]
1988	Read Codes, version 2	Recommended as standard terminology for general practice [30,000 terms]
1990		NHS purchases Read Codes
1992	Read Codes, version 3	NHS Clinical Terms, version 3 [150,000 terms]
1993	SNOMED Version 3.0	Focus on disorders, morphology, living organisms and social context. Veterinary content re-integrated. [96,042 concepts]
1997		LOINC codes integrated into SNOMED
1998	SNOMED Version 3.5	
1999	Clinical Terms, version 3	Expansion of NHS Clinical Terms, version 3 [250,000 terms]
2000	SNOMED RT	Logic based approach for definition of SNOMED concept, axes became hierarchies, polyhierarchical approach to classification [128,030 concepts]
2002	SNOMED CT	Merger of SNOMED RT and NHS Clinical Terms, version 3 [325,857 concepts] Spanish edition, April 2002
2003		German edition, April 2003 SNOMED CT freely available within the USA
2007		Transfer of SNOMED CT to the newly formed International Health Terminology Standards Development Organization

SNOMED CT, the fourth major edition of SNOMED, was published in 2002 and its

enhancements included the introduction of qualifiers and handcrafted browser

hierarchies. The merger of SNOMED and NHS Clinical Terms between 1999 and 2001

to create SNOMED CT represented a true confluence of the two terminologies. All

concepts represented within the Read Codes are also represented within SNOMED CT.

Subsequent bi-annual releases of SNOMED CT have added content, identified

redundancies, retired "not elsewhere classified" and ambiguous concepts, and added

additional features that enhance both interface and search functionality of the

terminology. SNOMED CT has been translated into Spanish and German and translation

into Danish is in progress.

SNOMED CT has been freely available within the United States of America since 2003

through an agreement between the American College of Pathologists and the National

Library of Medicine. The transfer of SNOMED CT from The American College of

Pathologists and SNOMED International to the newly formed International Health

Terminology Standards Development Organization occurred in 2007.

SNOMED CT as a terminology for the representation of clinical care

The primary purpose of a contemporary clinical terminology is to support the

documentation of clinical care through the representation of [patient] conditions and

events. The terminology must enable the representation of concepts that relate to health

care within the electronic health record and other clinical information management

systems. The dominant uses for a clinical terminology in the twenty-first century

encompass data collection, data interchange, clinical decision support, clinical audit,

business operations, public health surveillance and research. SNOMED CT is the most

sophisticated clinical terminology currently available for the representation of health care and has the design [structure and function] to fulfill the desired requirements for a contemporary clinical terminology.

The desiderata for a contemporary clinical terminology[3, 4] form the foundation for the structure and function of SNOMED CT. The thirteen desiderata bring together the themes of vocabulary content, concept orientation, concept permanence, non-semantic concept identifiers, a polyhierarchical structure, formal definitions, the rejection of "not elsewhere classified" terms, multiple granularities, multiple consistent views, context representation, graceful evolution and recognized redundancy as a set of desired requirements for a contemporary clinical terminology. The current statuses of the desiderata for a clinical terminology that have been embedded within SNOMED CT are documented in Table 3.

SNOMED International modelers have also emphasized the importance of the three additional criteria of understandability, reproducibility and usefulness.[1] First, a SNOMED concept must be described and defined in a way that enables its use without reference to additional or inaccessible knowledge. Second, once the concept has been described and defined, all users must use the concept in the same way. Finally, the SNOMED concept must be useful for the representation of a clinical concept used for health care or in a health-related domain.

Since the release of SNOMED version 3.0, several studies[5-8] have confirmed that the SNOMED family of terminologies provides a more comprehensive and consistent representation for clinical care than other classifications and terminologies.

Table 3. Desiderata for a contemporary clinical terminology and SNOMED CT

Desiderata	Status within SNOMED CT
Vocabulary content [ensure domain completeness] [maintain systematic approach to the updating of vocabulary]	SNOMED CT is more complete than any other terminology. There are requirements for additional contents, however, the design and structure of SNOMED CT support growth by many orders of magnitude if and as required.
Concept orientation [stress structural and knowledge representation issues — not just on expanding content]	The SNOMED CT Concept is the fundamental element of SNOMED CT.
Concept permanence [ensure concept permanence: do not delete a concept or change its meaning]	Once allocated a SNOMED CT Concept id is permanent.
Meaningless identifiers [do not include meaningless concept identifiers]	SNOMED CT identifiers are unique, semantically meaningless identifiers that do not attempt to convey semantic meaning. The functionality of SNOMED CT identifiers is quite different from the identifiers of ICD-9, the Read codes and other terminologies.
Polyhierarchy [specify multiple hierarchies rather than lists]	A SNOMED CT Concept can have any number of supertypes and subtypes. SNOMED CT supports multiple parent-child relationships within a vertical hierarchical structure.
Formal definitions [maintain formal definitions — replace informal descriptions]	SNOMED CT Concepts are described in terms of one or more defining relationships with other SNOMED CT concepts. Attributes are the fundamental component of the defining relationships. SNOMED CT concepts are categorized as primitive or fully defined.

Rejection of "not elsewhere classified" terms [eliminate "not elsewhere classified" terms which may be introduced to deal with incompleteness in a vocabulary]	SNOMED CT discourages the use of "not elsewhere classified" terms. Not Elsewhere Classified concepts are included only when essential for mapping to other terminologies or classifications. These concepts are marked with Concept Status 'limited" to discourage their use.
Multiple granularities [support multiple granularities to meet the differing needs of different users]	The SNOMED CT structure supports any number of levels of refinement. In most clinical areas useful concepts have subtypes, or can be qualified, to allow additional concepts to be added.
Multiple consistent views [maintain multiple consistent views: ensure consistency over different views of a hierarchy]	SNOMED CT subsets, navigation hierarchies and other interface support files assist the delivery of multiple configurable views tailored for particular disciplines, specialties and languages. These views do not interfere with the consistency of SNOMED CT as a whole.
Context representation [represent context specific information to maintain the relationship between a concept and the context in which it is used]	The SNOMED CT context model specifies a formal representation of context as it applies to SNOMED CT concepts and expressions.
Graceful evolution [allow for graceful evolution in vocabulary design to support inclusion of new developments in health care and error correction]	Six monthly releases of SNOMED CT supported by component history, historical relationships and references allow for the graceful addition to SNOMED CT content and the correction of SNOMED CT content.
Recognize redundancy [able to recognize redundancy where the same information is expressed in different ways]	Redundant SNOMED CT concepts can be recognized during the definition process. Redundant concepts are managed using the Concept Status "duplicate" and the "SAME AS" historical relationship.

Chute and colleagues[5] evaluated the content coverage of SNOMED, the Read Codes and

five other coding schemes. Table 4 lists the seven coding schemes that were evaluated.

The representation of 3,061 distinct concepts within the seven coding schemes was

determined. The concepts were categorized as Diagnoses, Modifiers, Findings,

Treatments and Procedures, and Other. Representation was assessed as absent, partial or

complete [scored as 0, 1 or 2 respectively].

Table 4. Coding Schemes evaluated by Chute and colleagues[5]

Coding Scheme	Full name of the Coding Scheme
CPT	Current Procedural Terminology, 1993 edition
ICD-9-CM	International Classification of Diseases, Ninth Revision, Clinical Modification, 1993 edition
ICD-10	International Classification of Diseases, 10th edition
NANDA	North American Nursing Diagnosis Association
Read	Read Coding Scheme, version 2
SNOMED III	SNOMED International, version 3.0
UMLS	Unified Medical Language System, version 1.3

SNOMED III had the highest overall score (1.74 out of 2.00) and the highest score in

every category including Diagnosis (1.90). Campbell and colleagues[6] subsequently

reported the evaluation of SNOMED III, UMLS version 1.6 and the Read codes version

3.1 for completeness, taxonomy, mapping, definitions and clarity. SNOMED III had the

highest overall score (4.31 out of 5.00) of the three coding schemes and had the richest

clinical taxonomy as judged by the number of acceptable first degree relatives per coded

concept. SNOMED III was judged to have the least clarity of the three schemes because

it had the most duplications of coding assignments. The Health Data Standards

Committee, which includes members from the National Library of Medicine and the

Agency for Health Care Policy and Research, evaluated the coverage of more than 30

controlled health data terminologies in the 1996 UMLS Knowledge Source with a large

vocabulary set of 41,127 terms.[7] Only SNOMED III and the Read Codes had more than

60% of the terms and more than 50% of the concepts.

An analysis of the adequacy of evolving national standardized terminologies for

interdisciplinary coded concepts in an automated clinical pathway also confirmed that

SNOMED CT provides very good representation for concepts that describe clinical care.[8]

Over 90% of the concepts were represented by one or more of the coding systems but

SNOMED CT and the Logical Observation Identifier Names and Codes (LOINC)

represented concepts better than five mandated terminologies. SNOMED CT and LOINC

alone represented 86.2% of the concepts but did not include representation for all

concepts necessary for complete electronic monitoring of guideline adherence.

Several studies have assessed SNOMED for the representation of the clinical problem

list[9-11] and reasons for visits to the Emergency Room.[12] SNOMED CT provided complete

representation for 88.4% and partial representation for 10.1% of a clinical vocabulary for

the computerized diagnosis and problem list.[9] Only 20 terms (1.5%) were not represented

by SNOMED CT. There was greater than 90% complete and partial representation for

Veterans Health Administration problem list entries.[10] These entries had a broad scope

that included symptoms, laboratory findings, diagnoses, procedures, medical risk terms,

information about patient participation in research trials, abbreviations and other terms.

Limitations in representation by SNOMED CT were attributed to the use of abbreviations

13

and spelling errors. Elkin and colleagues[11] evaluated the use of SNOMED CT as a compositional terminology for the representation of 4996 clinical problems seen at the Mayo Clinic, Rochester, Minnesota. There was complete representation for 4568 (92.3%) of the clinical problems. A post-coordinated SNOMED CT compositional expression was required for 2021 (51.4%) of the 4568 problems to achieve complete representation. The investigators considered that improvements to synonymy and the addition of missing modifiers would improve the representation of common clinical problem statements by SNOMED CT. McClay and Campbell[12] judged SNOMED CT to provide complete representation for 92.9% of the 296 reasons for visits to the Emergency Room at a 700 bed tertiary care hospital in Nebraska. There was partial representation for 18 reasons (6.1%) and no representation for 3 reasons (1.0%). The reasons for partial representation of 18 reasons for visits were different concepts (4), SNOMED concepts too narrow (9) and SNOMED concepts too broad (5).

Min and colleagues[13] mapped 689 unique terms for cardiovascular disease from OpenSDE domain models to SNOMED CT. The terminology represented 647 terms (93.9%). There was complete representation for 616 terms (89.4%) and partial representation for 31 terms (4.5%). Brown and colleagues[14] evaluated the ability of SNOMED CT to represent Federal Drug Agency approved oncology drug indications. The terminology represented 1317 of the 1527 (86.3%) single concepts present in the drug indications. Only 13 indications (11.3%) had all their concepts represented. Representation was best for concepts describing diseases, anatomy and patient characteristics. Medications accounted for 50.5% of the missing concepts. Representation increased to 91.2% if 340 names of drugs were excluded and there was complete

representation for 45.2% of the oncology drug indications. Li and colleagues[15] mapped a

set of 1044 pathogen codes from the National Nosocomial Infection Surveillance (NNIS)

system to SNOMED concepts. There was representation for 987 (94.5%) of the NNIS

concepts by SNOMED concepts. Reversal of the direction of the mapping resulted in

representation for only 7.6% of the 12,960 SNOMED infectious agents by a NNIS code.

SNOMED CT is a terminology that may be used for the representation not only of

clinical care but clinical research. Richardson and colleagues[16] assessed the

representation provided by SNOMED CT for a set of 616 unique clinical research

concepts extracted from the case report forms of observational studies of vasculitis. Most

of the core clinical concepts (88%) were represented by SNOMED CT. Just over 75% of

the concepts required post-coordinated SNOMED CT compositional expressions to

achieve full representation. These compositional expressions were necessary for the

capture of complex clinical concepts or the clarification of context. Both types of

compositional expression were necessary for complete representation of the clinical

research concept for just over 33% of the concepts. Although SNOMED CT represented

a variety of clinical concepts very well, it was judged to be less suitable for the

representation of the full amount of information generated by participation in clinical

research.

SNOMED CT is an established terminology for the expression of human and veterinary

medical concepts. Zimmermann and colleagues[17] judged SNOMED representation of

explanatory knowledge in veterinary clinical pathology to be fair. Veterinary pathology

concepts were extracted from three types of clinical pathology documents — a textbook,

the results and discussion sections of industrial pathology reports. There was

15

representation by SNOMED CT concepts for 442 (86.2%) of the 492 veterinary

pathology terms. Only 233 (47.4%) of the terms were completely represented while

38.8% were partially represented. SNOMED CT representation was absent for 68

(13.8%) of the veterinary pathology terms. The investigators observed that the missing

and problem concepts were confined to a relatively small area of the terminology. They

commented that if future revisions of SNOMED CT addressed this limitation, the

relevance and application of SNOMED CT to veterinary clinical pathology would be

enhanced. The evaluation of highly specialized medical domains might yield similar

results and observations.

An evaluation of SNOMED 3.5 for the representation of terms in chest radiology reports

also confirmed that SNOMED was less suitable for the representation of highly

specialized clinical domains.[18] SNOMED had corresponding concepts for the

representation for 72% of 700 anatomical location terms. The remaining 28% of the

terms were judged to be plural terms (4%), morphological variants of SNOMED 3.5

concepts (7%), terms that required a post-coordinated SNOMED compositional

expression for full representation (5%) or terms that were not included in SNOMED 3.5

(12%). There was complete representation for 77% of the radiologic terms together with

partial representation for 11% of the radiological terms. Only 12% of the terms, which

were predominantly specialized radiologic terms, were concepts that were not

represented by SNOMED 3.5 concepts.

Osteopathic concepts were very poorly represented by SNOMED CT.[19] Only 15% of 671

terms extracted from the Authorized Osteopathic Thesaurus of the American Association

of Colleges of Osteopathic Medicine were represented by SNOMED CT concepts. Many

commonly used terms that describe assessment and physical examination findings were missing from SNOMED CT. The majority of common osteopathic procedures were not represented within the terminology. The integration of essential osteopathic concepts into SNOMED CT will lead to an improved ability to capture, retrieve and aggregate health care data for osteopathic patients.

Ophthalmology is another highly specialized medical domain where the representation of unique concepts within clinical information management systems is important. Chiang and colleagues[20] have evaluated the adequacy of controlled medical terminologies for the representation of ophthalmology concepts. Lee and colleagues[21] have evaluated SNOMED CT for the representation of aesthetic concepts commonly encountered in the oculoplastic and facial plastic surgery domains. The adequacy of ICD-9-CM, CPT-4, SNOMED CT, LOINC and the Medical Entities Dictionary (MED) were evaluated for their representation of 1603 concepts extracted from 20 complete ophthalmology case presentations sequentially selected from a publicly available ophthalmology journal. SNOMED CT was judged to have the best performance with a mean overall coverage score of 1.625 ± 0.667 (maximum score 2.00). The scores for MED and LOINC were between 0.75 and 1.00 while those for ICD9-CM and CPT-4 were very low (Table 5). SNOMED CT had the highest coverage scores for concepts in the diagnosis, finding and procedure categories. Furthermore, general medical concepts were better represented than specialized ophthalmologic concepts. Lee and colleagues[21] established that the representation of 62 aesthetic concepts commonly encountered in the oculoplastic and facial plastic surgery domains was very poor. SNOMED CT was unable to represent 68% of these concepts. There was partial representation for 13% of the concepts together with

commonly used terms that describe assessment and physical examination findings were missing from SNOMED CT. The majority of common osteopathic procedures were not represented within the terminology. The integration of essential osteopathic concepts into SNOMED CT will lead to an improved ability to capture, retrieve and aggregate health care data for osteopathic patients.

Ophthalmology is another highly specialized medical domain where the representation of unique concepts within clinical information management systems is important. Chiang and colleagues[20] have evaluated the adequacy of controlled medical terminologies for the representation of ophthalmology concepts. Lee and colleagues[21] have evaluated SNOMED CT for the representation of aesthetic concepts commonly encountered in the oculoplastic and facial plastic surgery domains. The adequacy of ICD-9-CM, CPT-4, SNOMED CT, LOINC and the Medical Entities Dictionary (MED) were evaluated for their representation of 1603 concepts extracted from 20 complete ophthalmology case presentations sequentially selected from a publicly available ophthalmology journal. SNOMED CT was judged to have the best performance with a mean overall coverage score of 1.625 ± 0.667 (maximum score 2.00). The scores for MED and LOINC were between 0.75 and 1.00 while those for ICD9-CM and CPT-4 were very low (Table 5). SNOMED CT had the highest coverage scores for concepts in the diagnosis, finding and procedure categories. Furthermore, general medical concepts were better represented than specialized ophthalmologic concepts. Lee and colleagues[21] established that the representation of 62 aesthetic concepts commonly encountered in the oculoplastic and facial plastic surgery domains was very poor. SNOMED CT was unable to represent 68% of these concepts. There was partial representation for 13% of the concepts together with

complete representation for 19% of the concepts. The reasons for partial representation of 8 of the concepts were different concepts (2), SNOMED concepts too narrow (2) and SNOMED concepts too broad (4).

Table 5. Representation of ophthalmology concepts [Chiang and colleagues[20]]

Controlled terminology	Mean coverage score	Standard deviation
SNOMED CT	1.625	0.667
MED	0.974	0.764
LOINC	0.781	0.929
CD9-CM	0.280	0.619
CPT-4	0.082	0.337

Evaluations of the performance of the SNOMED family of terminologies have confirmed that SNOMED provides comprehensive and consistent representation of the general medical concepts involved in clinical care.[5-13, 15] The performance of the SNOMED family of terminologies for the representation of concepts for clinical care is summarized in Table 6.

The degree of representation for health care concepts by SNOMED concepts falls as the medical domain becomes increasingly specialized. Previous studies have shown that representation is fair for oncology drug indications,[14] veterinary clinical pathology,[17] radiology [anatomical location][18] and ophthalmology concepts.[20] Representation is poor for osteopathic,[19] oculoplastic and facial plastic surgical concepts.[21]

Table 6. Representation of health care concepts by the SNOMED family of terminologies

Study	Domain	N	None	Partial	Complete
Li, 2003[15]	Microbiology	1044	5.5%	-	94.5%
Brown, 2003[14]	Oncology [drug indications]	1527	13.7%	11.3%	75.0%
McClay, 2002[12]	Emergency medicine	689	6.1%	4.5%	89.4%
Min, 2006[13]	Cardiovascular disease	296	1.0%	6.1%	92.9%
Wassermann, 2003[9]	General medical [problem]	1105	1.5%	10.1%	88.4%
Elkin, 2006[11]	General medical [problem]	4996	na	na	92.3%
Zimmermann, 2005[17]	Veterinary medicine	496	13.8%	38.8%	47.4%
Richardson, 2006[16]	Clinical research [vasculitis]	616	12%	20%	68%
Sinha, 2000[18]	Radiology	700	12%	11%	77%
Nash 2005[19]	Osteopathic medicine	671	85%	-	15%
Chiang, 2005[20]	Ophthalmology	1603	11.2%	15.2%	73.6%
Lee, 2006[21]	Ophthalmology	62	68%	13%	19%

SNOMED CT has evolved from the Systematized Nomenclature of Pathology [SNOP, 1965] into a Systematized Nomenclature of Medicine [SNOMED 1974], a reference terminology [SNOMED RT, 2001] and a clinical terminology [SNOMED CT, 2002] over forty years. A recent review of the published scientific literature about the SNOMED family of terminologies reveals a shift in emphasis from pathology to other medical specialties since 1990.[22] There has been a minor contribution to the SNOMED literature from pathology since the release of SNOMED CT. There is a striking absence of published scientific SNOMED literature for the child healthcare domains of pediatrics and neonatology. The content of SNOMED CT has been determined largely by the voluntary contributions of many, diverse clinical groups. An unforeseen consequence of

this opportunistic evolutionary process may be that some unique, clinically relevant concepts of highly specialized clinical domains are still not represented in SNOMED CT.

The purpose of this project was to assess the performance of SNOMED CT as the terminology for representation of clinical care of the newborn infant. The primary objective of this project was to determine whether commonly used terms that are relevant for the clinical care of the newborn infant are represented by SNOMED CT concepts.

Methods

The SNOMED CT July 2006 release was systematically examined using the CliniClue[®] 2006 Terminology Browser[23] available for this release to determine if 881 terms that are relevant for the clinical care of the newborn infant are represented by SNOMED CT concepts. The terms were categorized as diagnosis [disorder], procedure, observation [clinical finding] and drug [pharmaceutical/biologic product]. The terms for diagnosis, procedure and observation were extracted from the Clinical Information Management System (CIMS)[a] used in the Neonatal Intensive Care Unit at the Hospital for Sick Children, Toronto, Ontario, Canada. The current edition of the Hospital's Drug Handbook and Formulary[24] was the source for drug terms.

The Hospital for Sick Children is a health care, teaching and research centre dedicated exclusively to children that is affiliated with the University of Toronto. The Hospital's NICU is a 32 bed regional, referral unit that provides tertiary and quaternary services for critically ill newborns within the Central East health region of Ontario. The region has a population close to six million with nearly 60, 000 births annually. Approximately 750 -

_____

[a] Sunrise Critical Care (Emtek), Eclipsys [http://www.eclipsys.com]

800 neonates are admitted to the NICU each year. Sixty percent of admissions to the NICU are mature, full-term neonates. The patient population includes mature, critically ill neonates with disorders involving one or more organ systems, and preterm infants with the spectrum of clinical disorders associated with prematurity. The dominant disorders include hypoxic-ischemic encephalopathy secondary to perinatal asphyxia, respiratory disorders of the newborn [transient tachypnea of the newborn, meconium aspiration, pneumonia, respiratory distress syndrome, persistent pulmonary hypertension of the newborn and congenital abnormalities of the lung], congenital cardiovascular malformations, congenital malformations of the gastrointestinal system, complications of prematurity requiring surgical intervention [persistent patent ductus arteriosus, necrotizing enterocolitis, post-hemorrhagic hydrocephalus and retinopathy of prematurity]. Other disorders include convulsions due to a variety of causes, fulminant sepsis with multiorgan failure, meningitis, metabolic disorders, neonates with chromosomal abnormalities, syndromes and multiple congenital anomalies.

CIMS was implemented in 1999 and has been customized steadily with regular in-house upgrades in response to user defined needs. Structured data entry for diagnosis and intervention were added to CIMS in 2003. The Diagnosis List was developed to enable structured data entry of diagnosis for into CIMS for clinical purposes. The diagnosis is exported from CIMS into a relational database for administrative requirements and epidemiologic research. Terms for diagnosis that were currently used by the neonatologists were selected from ICD-9-CM and ICD-10. Some additional terms were added either to enhance the representation of the pathophysiology or severity of the

disorder, or to represent "new" clinical concepts used in this NICU. The additional terms

for diagnosis are listed in Table 7.

Table 7. Additional terms for diagnosis included in the CIMS Diagnosis List

Term for diagnosis	Comment
Congestive heart failure	2 new terms for representation of pathophysiology [low cardiac output syndrome, pulmonary overcirculation]
Hypotension	3 options for representation of severity [mild, moderate, severe]
Hypoxic-ischemic encephalopathy	3 stages for representation of severity [Sarnat stages 1-3]
Necrotizing enterocolitis	5 stages for representation of severity [Bell stages 1, 2A, 2B, 3A, 3B]
Patent ductus arteriosus	2 new terms for representation of severity and/or intervention [PDA—indomethacin, PDA—ligation]
Pericardial effusion	3 options for representation of severity [asymptomatic, symptomatic, cardiac tamponade]
Perinatal depression	New term [neonate depressed at birth but presence of asphyxia uncertain/doubtful]
Periventricular leucomalacia	2 options for representation of severity
Poor vascular access	New term
Retinopathy of prematurity	5 stages for representation of severity [ROP stages 1-5]
Subglottic edema	New term
Thrombus	11 terms for representation of anatomical location (e.g. aortic thrombus, portal vein thrombus, renal vein thrombus, sagittal sinus venous thrombus)

The Procedure List was a local compilation of surgical procedures currently performed on newborns receiving intensive care in the NICU. Terms in the Procedure List were not selected from either the Current Procedural terminology or the Canadian Classification of Interventions. Observations were the terms that had been used by nurses for data entry since the implementation of CIMS in 1999.

A CIMS term was identified a priori as a unique neonatal concept if the CIMS term included the word neonatal or the usage of the CIMS term was restricted to the neonatal period and post-priori if the relevant SNOMED CT concept included the word neonatal. The 881 CIMS terms were distributed among the four sets of terms as 172 observations, 434 diagnoses, 110 procedures and 165 drugs.

CliniClue 2006 [version 2006.2.8; October 7, 2006],[23] a look-up engine for SNOMED CT concepts, was used to systematically determine the representation of the 881 CIMS terms by SNOMED CT concepts. The CliniClue Terminology Browser has been developed by the Clinical Information Consultancy (www.clininfo.co.uk) and is provided as freeware without a stated or implied warrantee of fitness for purpose. SNOMED CT concepts, descriptions and relationships for the July 2006 release were provided by Dr. Kent Spackman, Scientific Director, SNOMED International, as a Microsoft Access 2000 file and three text files solely for the purpose of this Capstone Project.

SNOMED CT was interrogated in UK English using the Find function of CliniClue 2006. The predetermined search features included all languages, unfiltered Concept status or any active status, and all Description types. No restrictions were applied to the search. The Find function offers five search options — "Words – any order", "Phrase match",

"Identical term", "Starts with" and "Ends with". The option "Words – any order" was always used for the first search. If the initial search did not return a match, the other options were used sequentially in no particular order until a match was found or all options failed to return a match. Alternative but semantically identical words or phrases were used as the search term if the initial sequence of searches failed to return a match.

CliniClue 2006 displays Concept Id, Description Id, Hierarchy [Subtype hierarchy, Concepts with related value] and SNOMED CT Concept definition. Data displayed as SNOMED CT Concept definition included Concept Status, Descriptions with Fully Defined Name, Preferred term and synonyms, Definitions and Codes. Concept Id, Description Id, Concept Status, the fully specified name, the number of synonyms, the status of the Definition [primitive, fully defined] and the hierarchy tag were recorded in a Microsoft Excel spreadsheet for subsequent analysis. Table 8 displays the data that was collected during the interrogation of SNOMED CT with CliniClue 2006.

Table 8. Data collected during the interrogation of SNOMED CT

Fully specified name
Concept Id
Description Id
Concept Status [current or limited]
Status of the definition for the Concept [fully defined or primitive]
Hierarchy tag
Number of synonyms

The representation of each CIMS term within SNOMED CT was classified as complete, partial or absent (Table 9).

Complete representation was defined as the presence of a SNOMED CT concept for the CIMS term. CIMS terms that were completely represented within SNOMED CT were categorized as a simple or a compositional match to a SNOMED CT concept. A simple match requires a single, pre-coordinated SNOMED CT expression for complete representation of the CIMS term. A compositional match requires a post-coordinated SNOMED CT compositional expression with two or more SNOMED CT concepts to achieve complete representation of the CIMS term. Conformance with the SNOMED CT concept model in the formation of SNOMED compositional expressions was achieved by the use of only those attributes approved in the SNOMED concept model together with their appropriate domains and ranges.

Partial representation was defined as the presence of a related SNOMED CT concept for the CIMS term. The related SNOMED CT concept for each partially represented CIMS term was categorized as too broad, too narrow or different. CIMS terms that are partially represented by a SNOMED CT concept were examined to determine if a post-coordinated SNOMED CT compositional expression could be used achieve complete representation of the CIMS term. If complete representation of the CIMS term was possible with a post-coordinated compositional SNOMED expression, the term was judged to be completely represented with a compositional match.

Absent representation was defined as the absence of both a SNOMED CT concept and a related SNOMED CT concept for the CIMS term.

Table 9. Definitions for the representation of CIMS terms by SNOMED CT concepts

Representation of CIMS terms within SNOMED CT	Criterion for degree of representation
Complete representation - simple match - compositional match	Matching SNOMED CT concept - pre-coordinated SNOMED CT expression - post-coordinated SNOMED CT compositional expression
Partial representation	Related SNOMED CT® concept - related concept too broad - related concept too narrow - related concept different
Absent representation	No representation within SNOMED CT

Complete representation for terms from the drug set of CIMS terms was defined as an exact match for the name, preparation and concentration of the drug. The corresponding SNOMED CT concept was required to express all three elements for the concept to be judged as providing complete representation of the drug.

The data for all 881 CIMS terms and the four sets of terms was analyzed with simple descriptive statistics. The representation of the CIMS terms by SNOMED CT concepts is reported as the percentage of instances that meet the definition for complete, partial or absent representation.

Results

Eight hundred and fifty-one of the 881 CIMS terms (96.6%) are represented by SNOMED CT concepts. There is complete representation for 761 CIMS terms (86.4%) and partial representation for 90 CIMS terms (10.2%). Thirty CIMS terms (3.4%) are not represented by SNOMED CT concepts. The representation of all CIMS terms and those of the four sets of CIMS terms are summarized in Table 10.

Complete representation is present for 90.8% of CIMS terms for diagnosis and 89.5% of CIMS terms for observation. There is complete representation for 80.9% of CIMS terms within the procedure set of terms and 75.2% of CIMS terms within the drug set of terms.

Partial representation is present for 6.4% of CIMS terms for diagnosis and 6.4% of CIMS terms for observation. There is partial representation for 14.6% of CIMS terms within the procedure set of terms and 21.2% of CIMS terms within the drug set of terms.

Table 10. Representation of CIMS terms by SNOMED CT concepts

Category of CIMS term	Number of CIMS terms	No representation	Partial representation	Complete representation
Diagnosis	434	2.8%	6.4%	90.8%
Observation	172	4.1%	6.4%	89.5%
Procedure	110	4.6%	14.6%	80.9%
Drug	165	3.6%	21.2%	75.2%
All terms	881	3.4%	10.2%	86.4%

Fourteen of the 30 CIMS terms (46.7%) that are not represented by SNOMED CT concepts are within the diagnosis set of CIMS terms. There are 7 observations, 6 drugs and 5 procedures that are not represented by SNOMED CT concepts.

Complete representation is achieved with a single, pre-coordinated SNOMED CT expression for 734 of the 761 CIMS terms (96.4%) that have complete representation within SNOMED CT. Twenty-five of the 757 CIMS terms (3.3%) require a post-coordinated SNOMED CT compositional expression for complete representation within SNOMED CT and ten of these require the use of an unapproved attribute (stage or grade). The 25 CIMS terms that require a post-coordinated SNOMED CT compositional expression for complete representation are listed in Table 11.

Seven hundred and eight of the 761 SNOMED CT concepts (93.0%) that completely represented CIMS terms have the current Concept Status. The 53 CIMS terms (7.0%) represented by SNOMED CT concepts with limited Concept Status are listed in Table 12. Fifty-two of these terms contain the word *other* and/or *unspecified.* These terms almost certainly have a common origin in classification-style coding systems such as ICD.

Two hundred and fifteen of the 761 SNOMED CT concepts (28.2%) are fully defined concepts. Three hundred and seventy six of these SNOMED CT concepts (49.4%) have one or more synonyms. The number of synonyms ranges from one to fourteen.

Table 11. Post-coordinated SNOMED CT compositional expression for CIMS terms

CIMS term	SNOMED CT compositional expression
Small bowel resection with plication	[small bowel resection] + [intestinal plication]
Enterotomy, duodenal tube	[incision and drainage of duodenum] + [insertion of nasoduodenal tube]
Enterotomy, jejunal tube	[incision and drainage of duodenum] + [insertion of nasoduodenal tube]
Flail chest, left	[flail chest] : [laterality = left]
Flail chest, right	[flail chest] : [laterality = right]
Hypotension, mild	[neonatal hypotension] : [severity = mild]
Hypotension, moderate	[neonatal hypotension] : [severity = moderate]
Hypotension, severe	[neonatal hypotension] : [severity = severe]
Ladd's procedure and fixation	[Ladd procedure for correction of volvulus] + [fixation of small intestine]
Malrotation with volvulus	[congenital malrotation of intestine] + [intestinal volvulus]
Necrotizing enterocolitis, stage 1*	[neonatal necrotizing enterocolitis] : [grade = 1]
Necrotizing enterocolitis, stage 2A*	[neonatal necrotizing enterocolitis] : [grade = 2A]
Necrotizing enterocolitis, stage 2B*	[neonatal necrotizing enterocolitis] : [grade = 2B]
Necrotizing enterocolitis, stage 3A*	[neonatal necrotizing enterocolitis] : [grade = 3A]
Necrotizing enterocolitis, stage 3B*	[neonatal necrotizing enterocolitis] : [grade = 3B]
Pleural effusion, asymptomatic	[pleural effusion] : [symptom = mild]
Pleural effusion, symptomatic	[pleural effusion] : [symptom = moderate]
Pulmonary venous embolism	[venous air embolism] : [finding-site = entire pulmonary vein]
Retinopathy of prematurity, stage 1*	[retinopathy of prematurity] : [stage = 1]
Retinopathy of prematurity, stage 2*	[retinopathy of prematurity] : [stage = 2]
Retinopathy of prematurity, stage 3*	[retinopathy of prematurity] : [stage = 3]
Retinopathy of prematurity, stage 4*	[retinopathy of prematurity] : [stage = 4]
Retinopathy of prematurity, stage 5*	[retinopathy of prematurity] : [stage = 5]
Transverse venous sinus thrombosis, left	[thrombosis transverse sinus] : [laterality = left]
Transverse venous sinus thrombosis, right	[thrombosis transverse sinus] : [laterality = right]

* Post-coordinated SNOMED compositional expression uses an unapproved attribute

Table 12. CIMS terms represented by SNOMED CT concepts with limited status

Anemia, unspecified
Asphyxia, unspecified
Bacteremia, other and unspecified staphylococci
Bacteremia, other and unspecified streptococci
Bacterial pneumonia, unspecified
Bacterial sepsis of newborn, unspecified
Birth injuries to other parts of peripheral nervous system
Birth injuries to other parts of skeleton
Birth injury to central nervous system, unspecified
Birth injury to scalp, unspecified
Birth injury to skeleton, unspecified
Birth injury, unspecified
Congenital viral disease, unspecified
Disorders of muscle tone of newborn, unspecified
Hydrops fetalis due to other and unspecified hemolytic disease
Ileus, unspecified
Jaundice due to excessive hemolysis, unspecified
Jaundice due to other specified excessive hemolysis
Jaundice from other and unspecified hepatocellular injury
Jaundice from other specified causes
Jaundice, unspecified
Neonatal jaundice due to other excessive hemolysis
Neonatal tetany without calcium or magnesium deficiency
Other bacterial meningitis
Other bacterial pneumonia
Other bacterial sepsis of newborn
Other birth injuries
Other birth injuries to central nervous system
Other birth injuries to scalp (includes caput succedaneum)
Other birth injuries to skull
Other congenital infections and parasitic diseases
Other congenital infectious and parasitic diseases, unspecified
Other congenital viral diseases
Other disorders of muscle tone of newborn
Other fungal meningitis
Other hemolytic diseases of fetus and newborn
Other infection of the newborn, unspecified
Other meningitis

Other neonatal hypocalcaemia
Other neonatal hypoglycemia
Other specified birth injuries
Other specified brain damage due to birth injury
Other specified congenital infectious and parasitic diseases
Other transient disorders of calcium and magnesium metabolism
Other transient disorders of calcium and magnesium, unspecified
Other transient disorders of carbohydrate metabolism
Other transient disorders of carbohydrate metabolism, unspecified
Other transient electrolyte disturbances of newborn
Other viral pneumonia
Thrombocytopenia, unspecified
Transient metabolic disturbances of newborn, unspecified
Unspecified brain damage due to brain injury
Viral pneumonia, unspecified

Two hundred and forty-four of the 761 SNOMED CT concepts (32.1%) represented CIMS terms that were judged to be unique neonatal concepts by a single clinician. The majority (95.5%) of these terms are from the Diagnosis set of CIMS terms. Fifty-four percent of the terms in the diagnosis set are unique neonatal concepts. The number of unique neonatal concepts in the Observation, Procedure and Drug sets of CIMS terms are 9 (5.8%), two (1.8%) and zero respectively. Properties of the SNOMED CT concepts that provide complete representation for CIMS terms are summarized in Table 13.

Ninety of the 881 CIMS terms (10.2%) were partially represented by related SNOMED concepts (Table 14). The reason for partial representation was judged to be a semantic difference between the CIMS term and the related SNOMED concept (5), the related SNOMED concept was too narrow (24) or the related SNOMED concept was too broad (71). The reason for partial representation for all CIMS terms in the drug set was the related SNOMED concept was too broad.

Table 13. Properties of SNOMED CT concepts that provide complete representation for CIMS terms

Category of CIMS term	Number of terms	Fully defined concept	Current Status	Pre-coordinated SNOMED concept	Synonyms present
Diagnosis	394	36.5%	86.5%	96.2%	45.4%
Observation	154	10.4%	100.0%	98.1%	61.0%
Procedure	89	61.8%	100.0%	95.5%	65.2%
Drug	124	0%	100.0%	100.0%	36.3%
All concepts	761	28.2%	93.0%	97.4%	49.4%

Table 14. Reason for partial representation of CIMS terms by related SNOMED concepts

Category of CIMS term	Number of terms	Semantic difference	SNOMED concept too narrow	SNOMED concept too broad
Diagnosis	28	10.7%	57.1%	32.1%
Observation	11	18.2%	27.3%	55.4%
Procedure	16	0%	31.2%	68.7%
Drug	35	0%	0%	100%
All terms	90	5.6%	26.7%	78.9%

The 30 CIMS terms (3.4%) that are not represented by SNOMED CT concepts are listed by category of term in Table 15. Six of the 12 CIMS terms for diagnosis contain the word *unspecified.* Four of the six drug terms are unique formulations prepared by the Hospital's pharmacy for administration exclusively within The Hospital for Sick Children, Toronto.

Table 15. CIMS terms that are not represented by SNOMED CT concepts

Diagnosis
Birth injury to peripheral nervous system, unspecified
Craniofacial
Hemolytic diseases of fetus and newborn, unspecified
Non-hemolytic jaundice
Other neonatal stroke, unspecified
Perinatal depression
Suboptimal vascular access
Transient disorders of calcium and magnesium metabolism
Unique perinatal diagnosis
Unspecified disturbance of potassium balance of newborn
Unspecified disturbance of sodium balance of newborn
Unspecified intrauterine growth restriction
Observation
High pitch bowel sounds
Low pitch bowel sounds
Migration irregular [seizure]
Migration ordered [seizure]
Not tested
Pulls away
Unable to assess
Procedure
Bowel lengthening
Intraluminal sling, intestinal
Intraluminal sling, esophageal
Rectal translocation
Stricture bypass
Drug
Dinoprostone [solution]
Epinephrine injection preloaded 1mg/10 ml
Erythromycin lactobionate [50mg/ml]
Flucytosine suspension [10mg/ml]
Mannitol 25%
Pyridoxine [solution 1mg/ml]

Discussion

SNOMED CT is the most sophisticated clinical terminology currently available for the representation of health care and has the ability to fulfill the desired requirements for a clinical terminology because the thirteen desiderata for a clinical terminology of the twenty-first century[3, 4] have been integrated into the design of SNOMED CT. The additional criteria of understandability, reproducibility and usefulness have also enhanced the performance of the terminology. The comprehensiveness, diversity and granularity of SNOMED CT has increased as additional concepts have been incorporated into the terminology from classifications and other terminologies during its graceful evolution over the past forty years. SNOMED CT contains more than 370,000 active concepts, more than one million active descriptions and nearly 1.5 million relationships.[2] Increasingly, SNOMED CT is being adopted in many countries as the preferred clinical terminology for the documentation of clinical care.

Many studies[5-13, 15] have confirmed that the SNOMED family of terminologies provides a more comprehensive and consistent representation for clinical care than other classifications and terminologies. The SNOMED family of terminologies provides very good representation for general medical[5-12] and selected specialized medical domains.[13, 20] The SNOMED CT July 2006 release provides very good representation for CIMS terms that represent the clinical care of the newborn. Over 96% of the terms are represented by SNOMED CT concepts although 10% of the terms are only partially represented. Less than 4% of the terms are not represented by SNOMED CT concepts. Terms in the categories of diagnosis and observation have the highest representation by SNOMED CT concepts. The CIMS terms for diagnosis were derived from ICD-9-CM and ICD-10,

which have already been integrated into SNOMED CT. The CIMS terms for observation are terms typically used by nursing. NANDA, NIC, NOC, NANDA, PNDS, the Clinical Care Classification and The Omaha System have also been integrated into SNOMED CT.[2] Although Neonatology is a specialized domain of medicine with many disorders that only occur in the neonatal period, the domain represents the health care for a specific, well-defined population as distinct from other specialties that represent health care for those with disorders related to a specific organ system. The majority of the terms that represent the care of the newborn have been derived from the domains of general medicine and surgery. Only one third of the CIMS terms were judged to be unique neonatal concepts by a single clinician with the majority of those terms describing diagnosis. The number of unique neonatal concepts in the Observation, Procedure and Drug sets of CIMS terms is negligible. The representation of CIMS terms for the care of newborns by SNOMED CT concepts is very good and is consistent with those for other domains that have their origins within general medicine and surgery.[5-13, 15, 20]

The majority of the CIMS terms (97.4%) that are completely represented by SNOMED CT concepts are represented by a single, pre-coordinated SNOMED CT expression. All CIMS terms in the Drug category and more than 98% of terms in the Observation category are represented by a pre-coordinated SNOMED CT expression. More than 95% of CIMS terms in the Diagnosis and Procedure categories are represented by a pre-coordinated SNOMED CT expression. The 25 CIMS terms that required a post-coordinated SNOMED CT compositional expression for complete representation are terms that demand greater granularity than those that do not require post-coordination. Ten of the compositional expressions use the unapproved attributes of stage and grade to

achieve complete representation of the severity of the disorder. The CIMS terms

necrotizing enterocolitis and retinopathy of prematurity have multiple stages that are fully

expressed only by a post-coordinated SNOMED CT compositional expression. Staging

and grading are common requirements for the representation of health care that are not

yet structured in a generalize manner by the approved SNOMED concept model. Some

would argue that stage 3A necrotizing enterocolitis and Stage 2 retinopathy of

prematurity are best represented as subtypes of the supertypes necrotizing enterocolitis

and retinopathy of prematurity. Post-coordination SNOMED CT compositional

expressions are also required to represent the severity of a clinical finding [hypotension,

pleural effusion] or the laterality of a clinical finding [flail chest, transverse sinus venous

thrombosis]. The documentation that less than three percent of 761 completely

represented CIMS terms require post-coordinated SNOMED CT compositional

expression supports the observation that SNOMED CT is a rich, comprehensive and

granular clinical terminology.

The majority of the SNOMED CT concepts (93%) that completely represented CIMS

terms have a current Concept Status. All CIMS terms in the Observation, Procedure and

Drug categories have a current Concept Status. The 53 CIMS terms (13.5%) that do not

have an active Concept Status reside within the Diagnosis set. Examination of the 53

terms that are represented by a SNOMED CT concept with a limited Concept Status

revealed that 52 terms contain the word *other* and/or *unspecified*. The other term is

"Neonatal tetany without calcium or magnesium deficiency". The interpretation for this

term is an instance of tetany in a newborn infant who has normal serum calcium and

magnesium levels. The term is meaningful as a clinical finding but provides less meaning

if used to represent a disorder. None of these three expressions should be included in a contemporary clinical terminology because they do not conform to the accepted desiderata for a contemporary clinical terminology. Many of these SNOMED CT concepts are legacy concepts inherited from classifications or terminologies that have been integrated into the SNOMED family of terminologies. SNOMED CT has incorporated the desideratum for the rejection of "not elsewhere classified" terms and discourages the use of "not elsewhere classified" terms. Not Elsewhere Classified concepts are included only when essential for mapping to other terminologies or classifications. These concepts are marked with Concept Status 'limited" to discourage their use. SNOMED concepts with limited Concept Status can be expected to become completely inactive concepts with the continuing graceful evolution of SNOMED CT.

The majority of concepts within a contemporary clinical terminology should be fully defined to enable and enhance the computational processes of auto-classification and the computation of equivalence and subsumption between different methods of expressing the same meaning. Nearly 30% of the completely represented CIMS terms are represented by fully defined SNOMED CT concepts. The highest number of fully defined SNOMED CT concepts occurs for the representation of procedures (61.8%) and diagnosis (36.5%). Just over 10% of terms in the Observation category are represented by fully defined SNOMED CT concepts while none of the terms in the Drug category are fully defined. Many SNOMED concepts are not fully defined at the present time because they are very general or non-clinical concepts, or the attributes or values necessary for the full definition of the concept are missing. Furthermore, it takes time to fully define over 370,000 current SNOMED concepts in a large terminology that has over one million

descriptions and nearly 1.5 million relationships. The attainment of fully defined status for more than 70% of SNOMED CT concepts is one of the stated goals for the terminology.

Nearly 50% of the SNOMED concepts that completely represent CIMS terms have synonyms. More CIMS terms in the Observation and Procedure categories are represented by SNOMED CT concepts with one or more synonyms than terms in the Diagnosis and Drug categories. The existence of one or more synonyms enhances the terminology's ability to function as an interface terminology, and therefore, should facilitate the accurate entry of valid data.

The continuing graceful evolution of SNOMED CT can be expected to increase the number of fully defined SNOMED CT concepts together with an increased number of SNOMED CT concepts that have one or more synonyms. A continuous increase in the number of fully defined concepts and the number of concepts with one or more synonyms will enhance the functionality of the terminology.

The dominant reason for partial representation of the 90 CIMS terms is the related SNOMED concept was too broad. All the related SNOMED concepts that partially represented drug terms are too broad and more than 50% of the related SNOMED concepts that partially represented observation and procedure terms are too broad. One third of the related SNOMED concepts that partially represented diagnosis terms are too broad. The related SNOMED concepts that are too narrow account for just over 25% of the instances of partial representation. Just over 5% of the instances of partial representation are due to a semantic difference between the CIMS term and the related

SNOMED concept. The addition of subtypes to selected SNOMED concepts will change the partial representation of many CIMS terms into complete representation. The representation of drug terms will improve with the addition of more concepts to the pharmaceutical/biologic product subtype hierarchy.

Only 30 CIMS terms (3.4%) are not represented by SNOMED CT concepts. Examination of the 30 terms reveals that half of the unrepresented terms for diagnosis contain the word unspecified. These terms should not exist within a contemporary clinical terminology. Several of the CIMS terms in the Observation category are poorly expressed clinical ideas that, in their present format, do not merit representation by SNOMED CT concepts. Unrepresented CIMS terms in the Procedure category may be terms that are unique to the domain of neonatology and therefore, are very unlikely to be represented by SNOMED CT concepts. Similarly, four of the five drugs not represented by SNOMED CT concepts are unique drug preparations that are prepared by The Hospital for Sick Children's pharmacy for use exclusively within the Hospital. Furthermore, drugs for the neonatal population might not be represented very well at present since the SNOMED CT pharmaceutical/biologic products hierarchy is based upon proprietary and generic sources from the United States and the United Kingdom. These sources may provide a less comprehensive representation for drugs used in the neonatal population than for drugs used in the adult population. Although there should be better representation of drugs administered to the newborn population with the continuing evolution of SNOMED CT, drugs prepared by a hospital's pharmacy for use exclusively within that institution are unlikely to be well represented by SNOMED CT concepts.

A limitation of this study is the lack of an assessment of the degree of representation of

the CIMS terms by SNOMED CT concepts by two or more reviewers. Nearly all previously reported studies[5-18, 20, 21] have used two or more reviewers together with adjudication by an additional, independent reviewer whenever there is a discrepancy in the assessments of the primary reviewers. The design for this study was limited to a single reviewer because of the circumstances of the OHSU distance learning graduate course in biomedical informatics. It is possible that the reviewer may have incorrectly assessed CIMS terms as completely represented by post-coordinated SNOMED CT compositional expressions or partially represented by SNOMED CT concepts. Furthermore, CIMS terms may have been assessed as having no representation by SNOMED CT concepts because the reviewer failed to locate the corresponding SNOMED CT concept within the terminology. If the reviewer made an error in the assessment of CIMS terms as being completely represented by post-coordinated SNOMED CT compositional expressions for 20 instances that were partial representation, the complete representation for the CIMS terms would fall from 86.4% to 84.1% but the overall representation would remain unchanged at 96.6%. Similarly, if the reviewer made an error in the assessment of CIMS terms as being partially represented by SNOMED CT concepts for 90 instances that lacked representation, the complete representation for the CIMS terms would remain unchanged at 86.4% but the overall representation would fall from 96.6% to 86.4%. The complete and overall representation of CIMS terms by SNOMED CT concepts after both types of error would be similar to that reported in other studies.[5-13, 20] The substantive findings of the study would not be affected by either of these errors, or the simultaneous occurrence of both types of error. Nevertheless, the preferred approach for the assessment of the degree of representation of

the CIMS terms by SNOMED CT concepts remains assessment by two or more reviewers together with adjudication by an additional, independent reviewer whenever there is a discrepancy in the assessments of the primary reviewers.

The structure of SNOMED CT enables the terminology to be used as a coding schema, a classification, a reference terminology and an interface terminology. Each SNOMED concept has a unique semantically meaningless numeric identifier that supports the terminology's computational capabilities. The hierarchical structure of SNOMED CT supports the function of classification of clinical concepts into groups or sets. SNOMED CT may also function as a reference terminology because it has a hierarchical structure with subtype hierarchies and defining relationships. A reference terminology has been defined by Spackman and colleagues[25] as "a set of concepts and relationships that provide a common reference point for comparison and aggregation of data about the entire health care process, recorded by multiple different individuals, systems, or institutions." A robust reference terminology specifies relationships between concepts, defines concepts in terms of their relationships and can be processed by computers. The design [structure and functionality] of SNOMED CT satisfies the necessary requirements for a reference terminology.

Additional themes incorporated into SNOMED CT include natural terms, component persistence with tracking history and support for the implementation of SNOMED CT in software applications. The design of SNOMED CT supports the inclusion of natural terms that clinicians use very well. Clinicians determine the meaning of the natural terms they use by the way they use those terms. SNOMED CT reflects the meanings of those natural terms. It is widely recognized that SNOMED CT coverage remains incomplete

41

and continues to evolve to fill identified gaps. All SNOMED CT components have permanent unique identifiers. The SNOMED CT history and reference tables track all changes that have been implemented within SNOMED CT. The implementation of SNOMED CT in software applications is enhanced by guidance based on a shared understanding of SNOMED CT design and experience.

The design [structure and functionality] of SNOMED CT facilitates valid and accurate, structured data entry and thus enables the terminology to function as an interface terminology. SNOMED CT supports data entry quality through text searches, navigation hierarchies, subsets and pick lists. The user's choice of term is supported by SNOMED CT concepts that have a preferred term, synonyms and homonyms as well as a fully specified name. SNOMED CT also provides multi-dialect and multilingual support for SNOMED CT concepts.

Summary and Conclusion

Ninety-six percent of the CIMS terms that are relevant for the clinical care of the newborn infant are represented by SNOMED CT concepts. Just over 86% of the terms are completely represented, 10% are partially represented and less than four percent of the CIMS terms are not represented within the terminology. The majority of the CIMS terms (97.4%) that are completely represented by SNOMED CT concepts are represented by a single, pre-coordinated SNOMED CT expression. Ninety-three percent of the SNOMED CT concepts that completely represent CIMS terms have a current Concept Status. Nearly 30% of these SNOMED CT concepts are fully defined SNOMED CT concepts and nearly 50% of these SNOMED CT concepts have synonyms.

SNOMED CT is a comprehensive reference terminology that provides robust, structured

representation based on description logic for the majority of the CIMS terms that are

relevant for the clinical care of the newborn infant. The continuing, graceful evolution of

SNOMED CT can be expected to improve the representation of the clinical care of the

newborn infant.

References

1. College of American Pathologists, SNOMED International. SNOMED Clinical Terms® User Guide. July 2006 Release. Northfield, IL, USA; 2006.

2. SNOMED Clinical Terms® (SNOMED CT®) Core Content for the July 2006 Release. Northfield, IL, USA; 2006. Available at www.snomed.org [accessed October 27, 2006].

3. Chute CG, Cohn SP, Campbell JR, for the ANSI Healthcare Informatics Standards Board Vocabulary Working Group and the Computer-based Patient Records Institute Working Group on Codes and Standards. A framework for comprehensive health terminology systems in the United States: Development guidelines, criteria for selection, and public policy implications. J Am Med Inform Assoc. 1998; 5:503-510.

4. Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. Methods Inf Med. 1998; 37:394-403.

5. Chute CG, Cohn SP, Campbell KE, Oliver DE, Campbell JR. The content coverage of clinical classifications. For The Computer-Based Patient Record Institute's Work Group on Codes & Structures. J Am Med Inform Assoc. 1996; 3:224-33.

6. Campbell JR, Carpenter P, Sneiderman C, Cohn S, Chute CG, Warren J. Phase II evaluation of clinical coding schemes: completeness, taxonomy, mapping, definitions, and clarity. CPRI Work Group on Codes and Structures. J Am Med Inform Assoc. 1997; 4:238-51.

7. 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.

8. Dykes PC, Currie LM, Camino JJ. Adequacy of evolving national standardized terminologies for interdisciplinary coded concepts in an automated clinical pathway. J Biomed Inform. 2003;36:313-25.

9. Wasserman H, Wang J. An applied evaluation of SNOMED CT as a clinical vocabulary for the computerized diagnosis and problem list. AMIA Annu Symp Proc. 2003:699-703.

10. Penz JF, Brown SH, Carter JS, Nguyen VN, Sims SA, Lincoln MJ. Evaluation of SNOMED coverage of Veterans Health Administration terms. Medinfo 2004; 11:540-4.

11. 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; 81:741-8.

12. McClay JC, Campbell J. Improved coding of the primary reason for visit to the emergency department using SNOMED. Proc AMIA Symp. 2002:499-503.

13. Min Z, Baofen D, Weeber M, van Ginneken AM. Mapping OpenSDE domain models to SNOMED CT. Applied to the domain of cardiovascular disease. Methods Inf Med. 2006; 45:4-9.

14. Brown SH, Bauer BA, Wahner-Roedler DL, Elkin PL. Coverage of oncology drug indication concepts and compositional semantics by SNOMED-CT. AMIA Annu Symp Proc. 2003:115-9.

15. Li W, Tolson J, Horan TC. Creating public health standard vocabularies: mapping a set of CDC's pathogen codes to SNOMED concepts. AMIA Annu Symp Proc. 2003:907.

16. Richesson RL, Andrews JE, Krischer JP. Use of SNOMED CT to represent clinical research data: a semantic characterization of data items on case report forms in vasculitis research. J Am Med Inform Assoc. 2006; 13:536-46.

17. Zimmerman KL, Wilcke JR, Robertson JL, Feldman BF, Kaur T, Rees LR, Spackman KA. SNOMED representation of explanatory knowledge in veterinary clinical pathology. Vet Clin Pathol. 2005; 34:7-16.

18. Sinha U, Yaghmai A, Thompson L, Dai B, Taira RK, Dionisio JD, Kangarloo H. Evaluation of SNOMED3.5 in representing concepts in chest radiology reports: integration of a SNOMED mapper with a radiology reporting workstation. Proc AMIA Symp. 2000:799-803.

19. Nash S. Integrating Osteopathic terminology into SNOMED CT. AMIA Annu Symp Proc. 2005;1061.

20. Chiang MF, Casper DS, Cimino JJ, Starren J. Representation of ophthalmology concepts by electronic systems: adequacy of controlled medical terminologies. Ophthalmology. 2005; 112:175-83.

21. Lee S, Tsirbas A, Goldberg RA, McCann JD. Standardized terminology for aesthetic ophthalmic plastic surgery. Ophthal Plast Reconstr Surg. 2006; 22:371-4.

22. Cornet R, de Keizer N. Forty years of SNOMED: a literature review. Proceedings of The First Semantic Mining Conference on SNOMED CT®. Copenhagen, Denmark. October 1-3, 2006;54-58.

23. CliniClue 2006 [version 2006.2.8, October 7, 2006]. Clinical Terminology Services from The Clinical Information Consultancy. Available at www.cliniclue.com [accessed November 6, 2006].

24. The Drugs and Therapeutic Committee, The Hospital for Sick Children. The 2006-2007 Drug Handbook and Formulary (24th edition).

25. Spackman KA, Campbell KE, Cote RA. SNOMED RT: A reference terminology for health care. In Masys DR. (ed): Proceedings of the 1997 AMIA Fall Symposium 1997. Philadelphia, Hanley & Belfus: 640-644

Appendix 1. Top-level Subtype Hierarchies of SNOMED CT [July 2006 release]

SNOMED CT concept	Brief description of concept representation
Body structure	Concepts include normal and abnormal anatomical structures. Normal anatomical structures can be used to specify the body site involved by a disease or procedure. Attribute — laterality [left, right, left and right, unilateral].
Clinical finding	Concepts represent the result of a clinical observation, assessment or judgment [history and symptoms, signs, laboratory results, imaging findings, diagnoses etc], and include both normal and abnormal clinical states. Contains the sub-hierarchy of Disease — modeled as an abnormal clinical finding. Multi-axial subtype hierarchies allow disease to be subtypes of other disorders as well as subtypes of findings. Attributes include finding site, associated morphology, associated with [after, due to, causative agent], has interpretation, interprets, episodicity, pathological process, course, has definitional manifestation, occurrence, severity, finding method, finding informer.
Context-dependent category	A situation with explicit context [clinical situation]. Semantic-context refers to the effect that placing a concept in a clinical situation has on the interpretation of that concept. Attributes include associated finding [clinical finding], associated procedure [procedure], finding context finding context value], procedure context [context value for action], temporal context [temporal context value], subject relationship context [person].
Environments and geographical locations	Types of environments as well as named locations such as regions, states and countries.
Event	Concepts that represent things that can occur or happen.
Linkage concept	Concepts that are used for linkage. Contains the sub-hierarchies link assertion [has reason, has explanation] and attributes that may be defining attributes [is_a, concept model attributes] or non-defining attributes [concept history, unapproved attribute].

48

Observable entity	Represent entities that can be measured or observed.
Organism	Organisms of significance in human and animal disease.
Pharmaceutical/biologic product	Drugs and other biological agents. This top-level hierarchy was introduced to make a clear distinction between drugs [Pharmaceutical/biologic product] products and their chemical constituents [Substance].
Physical force	Concepts representing physical forces that can have a role as mechanisms of injury.
Physical object	Concepts include natural and man-made objects.
Procedure	Represent activities performed in the provision of healthcare.
Qualifier value	Concepts as values for SNOMED CT attributes that are not contained elsewhere in SNOMED CT.
Record artifact	An entity that is created by a person or persons for the purpose of providing other people with information about events or states of affairs.
Social context	Social conditions and circumstances relevant to healthcare. These concepts represent social aspects that affect patient health and treatment. Content includes family status, economic status, ethnic and religious heritage, life style, and occupation.
Special concept	Contains inactive concept [ambiguous, duplicate, erroneous, moved elsewhere, outdated, reason not stated], namespace concept and navigational concept
Specimen	Entities that are usually obtained from a patient for examination or analysis.
Staging and scales	Contains sub-hierarchies that pertain to the naming of assessment scales [assessment scale] or tumor staging systems [tumor staging].
Substance	Concepts that can be used for recording active chemical constituents of drugs, allergens, adverse reactions, toxicity or poisoning information, and physician and nursing orders.

49

Appendix 2. Approved attributes of the SNOMED CT [July 2006 release]

SNOMED CT concept	Attribute [permissible value]
Body structure	Laterality [left, right, right and left, unilateral].
Clinical finding	Finding site [anatomical concepts, acquired body structure]
	Associated morphology [morphologically abnormal structure]
	Associated with [clinical finding, procedure, event, organism, substance, physical object, physical force, pharmaceutical/biologic product]
	Causative agent [organism, substance, physical object, physical force, pharmaceutical/biologic product]
	Due to [clinical finding, event]
	After [clinical finding, procedure]
	Severity [severities]
	Onset [gradual onset, sudden onset]
	Course [courses]
	Episodicity [episodicities]
	Interprets [observable entity, laboratory procedure, evaluation procedure]
	Has interpretation [findings values, results comments]
	Pathological process [pathological process]
	Has definitional manifestation [clinical finding]
	Occurrence [periods of life]
	Finding method [procedure]
	Finding informer [performer of method, provider of history other than subject (person), subject of record (person), subject of record or other provider of history (person)]
Context-dependent category	Associated finding [clinical finding]
	Finding context [finding context value]
	Associated procedure [procedure]
	Procedure context [context value for action]
	Temporal context [temporal context value]

	Subject relationship context [person]
Pharmaceutical/biologic product	Has active ingredient [substance]
	Has dose form [type of drug preparation]
Procedure	Procedure site – direct [anatomical concepts, acquired body structure]
	Procedure site – indirect [anatomical concepts, acquired body structure]
	Direct morphology [morphologically abnormal structure]
	Indirect morphology [morphologically abnormal structure]
	Method [action]
	Procedure device [device, physical force]
	Direct device [device]
	Indirect device [device]
	Using [device, physical force]
	Access instrument [endoscope, device]
	Approach [procedural approach]
	Direct substance [substance, pharmaceutical/biologic product]
	Priority [priorities]
	Has focus [clinical finding, procedure]
	Has intent [intents (nature of procedure values)]
	Recipient category [person, family, community, donor for medical or surgical procedure, group]
	Revision status [primary operation, revision-value, part of multistage procedure]
Specimen	Specimen procedure [procedure]
	Specimen source topography [anatomical concept]
	Specimen source morphology [morphologically abnormal structure]
	Specimen substance [substance]
	Specimen source identity [person, family, community, donor, device, environments]