Enabling Advanced EHR Analytics with Clinical Terminology

Session 233, Date of Session February 23, 2017

David Markwell, Head of Education, SNOMED International
Speaker Introduction

David Markwell, MB BS, LRCP, MRCS

Head of Education
SNOMED International
Conflict of Interest

David Markwell, MB BS, LRCP, MRCS

Is employed fulltime as at SNOMED International, the not-for-profit organization responsible for delivering and maintaining SNOMED CT. Other than this he has no real or apparent conflicts of interest to report.
Agenda

• Clinical Analytics and its Potential Benefits
• Barriers to Clinical Analytics
• Clinical Terminology and its Role in Meaning Based Analytics
• Analytics with SNOMED CT – What is Being Done
• Approaches to Use of SNOMED CT in Analytics
• Barriers to Delivery of the Theoretical Value
• Recommendations and Future Directions
Learning Objectives

• Recognize the potential value of using a logically defined clinical terminology to support effective clinical analytics
• Describe ways in which the features of SNOMED CT have been applied to analytics by vendors and healthcare providers
• Compare alternative approaches to enhancing clinical and business intelligence using a formally defined clinical terminology
• Debate the strengths and weaknesses of SNOMED CT as part of a clinical analytics solution
An Introduction of How Benefits Were Realized for the Value of Health IT

Treatment/Clinical

Enhanced clinical analytics enhances quality of treatment by enabling:

• Research into effective care pathways

• Decision support that assists clinical teams to follow effective pathways

Savings

Improved care delivery is cost-effective as it reduces adverse events and unnecessary repetition of costly interventions
Clinical Analytics

Discovery, interpretation and communication of meaningful patterns in clinical data

Enables a cycle of benefits for
- Patients and Clinicians
- Populations
- Researchers
Clinical Analytics

Big Data Analytics
- Epidemiology
- Population Health Needs Assessment
- Clinical Audit
- Clinical Research
- Large Clinical Trials
- Gathering Evidence for Best Practice

Small Data Analytics
- Patient Record Summary
- Point of Care Reporting
- Knowledge Linkage
- Applying Best Practice Guidelines
- Adverse Reaction Reduction
- Decision Support
Barriers to Clinical Analytics: Inconsistent Representation of Clinical Ideas

• Different words with similar meanings
  – Excision, removal, appendectomy

• Words with different meanings
  – Discharge (from hospital or wound?)

• Interrelated meanings
  – Appendicitis
    • Inflammation, disorder of intestine
  – Peritonitis
    • Inflammation, disorder of abdomen
Barriers to Clinical Analytics: Inconsistent Representation of Context

- Past history of appendicitis
- Appendectomy done in 2007
- Abdominal pain
  - Present, absent, resolved
- Diagnosis
  - Differential, provisional, suspected, possible, confirmed, excluded …
Clinical Terminology and Analytics

• Consistent approaches to representation of clinical ideas and context are required to address barriers to clinical analytics

• A well designed clinical terminology should enable data captured in different ways to be represented in semantically compatible and comparable forms

• But what constitutes a well-designed clinical terminology?
Desiderata for Controlled Medical Vocabularies in the Twenty-First Century

1. Vocabulary content
2. Concept orientation
3. Concept permanence
4. Non-semantic concept identifiers
5. Polyhierarchy
6. Formal definitions (description logic)
7. Rejection of "not elsewhere classified" terms
8. Multiple granularities
9. Multiple consistent views
10. Context representation
11. Graceful evolution
12. Recognized redundancy

Cimino JJ
SNOMED CT Design Requirements

- The “Desiderata” specified by Dr. Cimino in 1998
  - Multilingual – to support international use
  - Extensible – to meet national and local requirements
  - Practically implementable in software
  - International Not for Profit Ownership

*plus*

*and from 2007*
Concept Orientation

• One unique code (identifier) for each distinct clinical meaning (concept)
• Terms with the same meaning are represented as descriptions linked to the same concept
• A description linked to a concept may contain a term in any language or dialect
How a Concept Oriented Terminology Supports Effective Clinical Analytics

• Different user interface terms link to the same concept identifier
  – Data capture can be customized for
    • Different clinical workflows
    • Different language, dialect and user preferences
• Analytic queries only need to search for the concept identifier
  – They do not need to take account of the language or term used by during data entry
Polyhierarchy

- In a polyhierarchy each concept can be related to several parent concepts
- Each parent represents a more general concept

Example

- Laparoscopic appendectomy is a subtype of
  - Appendectomy and is a subtype of
  - Endoscopic operation
How a Polyhierarchy Supports Effective Clinical Analytics

• A polyhierarchy allows queries to select concepts and their subtypes even if concepts are in overlapping categories. For example
  – A query for “appendectomy” and its subtypes would return all the specific types of appendectomy (including “laparoscopic appendectomy”)
  – A query for “endoscopic operation” would return all specific types of endoscopic operations (including “laparoscopic appendectomy”)
• In a simple classification or monohierarchy, a code can only belong in one hierarchy branch, and thus can only be in one category
Formal Description Logic Definitions

• Description logic definitions formally represent the meaning of concepts

Examples
• Appendicitis is defined as
  ▪ Inflammation of the
  ▪ Appendix structure
• Appendectomy is defined as
  ▪ Excision of the
  ▪ Appendix structure
How Description Logic Definitions Support Effective Clinical Analytics

• By enabling Description Logic (DL) classifiers to make logical inferences to validate and enhance the subtype polyhierarchy
  – Improving the completeness and accuracy of subtype queries

• By allowing more specific queries that select concepts based on the values of particular defining attributes
Queries Based on Logic Definitions

Query: Disorders with finding site kidney

Query: Disorders with associated morphology benign neoplasm

Renal cyst
- Associated morphology: Cyst
- Finding site: Kidney

Benign tumor of kidney
- Associated morphology: Benign neoplasm
- Finding site: Kidney

Benign neoplasm of bladder
- Associated morphology: Benign neoplasm
- Finding site: Bladder

Abscess of lung
- Associated morphology: Abscess
- Finding site: Lung

Abscess of bladder
- Associated morphology: Abscess
- Finding site: Bladder

Renal abscess
- Associated morphology: Abscess
- Finding site: Kidney

Lung cyst
- Associated morphology: Cyst
- Finding site: Lung

Benign tumor of lung
- Associated morphology: Benign neoplasm
- Finding site: Lung
Context Representation

- SNOMED CT defines context including
  - Past history
  - Family history
  - Planned procedures

Example

- History of appendectomy
  - Is a subtype of Past history of procedure
  - With associated procedure Appendectomy
How Context Representation Supports Effective Clinical Analytics

• Different EHR systems represent context in different ways using record structures or qualifier codes
• Overlooking contextual information during analytics can cause highly significant errors
• A common way to represent context using the terminology has the potential to reduce errors and improve consistency
• Context specific structures can still be useful but binding these to a common reference representation improves interoperability
From Theory to Practice

- The potential benefits of enhanced clinical analytics are well understood.
- In theory, a clinical terminology with the features identified in Dr Cimino’s 1998 paper can deliver valuable enhancements to clinical analytics.
- However, some practical questions remain open:
  - To what extent are the features of SNOMED CT being used to improve analysis of electronic health records?
  - What are the barriers to fully realizing the potential benefits?
  - How should those barriers be addressed?
Data Analytics Guidance Development

• In 2014 IHTSDO* launched a project to develop and publish guidance on use of SNOMED CT in data analytics

• Stage 1: A request for information about usage of SNOMED CT in clinical analytics by:
  – National and regional health bodies
  – Healthcare institutions
  – Vendors
  – Researchers

• Stage 2: Development of guidance for data analytics with SNOMED CT

* At the time SNOMED International was known as IHTSDO
### Summary of Respondent Types

<table>
<thead>
<tr>
<th>Organization type</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>EHR vendors</td>
<td>7</td>
</tr>
<tr>
<td>National bodies and healthcare providers</td>
<td>4</td>
</tr>
<tr>
<td>Vendors of terminology related services</td>
<td>3</td>
</tr>
<tr>
<td>Knowledge publishers</td>
<td>2</td>
</tr>
<tr>
<td>Researchers</td>
<td>1</td>
</tr>
</tbody>
</table>

These figures currently exclude respondents to a more recent request for information related to clinical decision support.
<table>
<thead>
<tr>
<th>Type of Use</th>
<th>Respondents Reporting this Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clinical reporting and aggregation</td>
<td>11</td>
</tr>
<tr>
<td>Decision support</td>
<td>9</td>
</tr>
<tr>
<td>Cohort identification</td>
<td>5</td>
</tr>
<tr>
<td>Case-mix, management and billing</td>
<td>5</td>
</tr>
<tr>
<td>Epidemiology statistics and ICD mapping</td>
<td>3</td>
</tr>
<tr>
<td>Linkage to publications</td>
<td>2</td>
</tr>
<tr>
<td>Research</td>
<td>2</td>
</tr>
<tr>
<td>Data quality analysis</td>
<td>1</td>
</tr>
</tbody>
</table>
Preparing Data for Analytics with SNOMED CT

Data recorded as other codes

Data recorded as SNOMED CT

Data recorded as text

Mapping to SNOMED CT

Data for local SNOMED CT analysis

Natural Language Processing (NLP)

Data warehouse for aggregate SNOMED CT analysis
### Preparing Data for SNOMED CT Analytics - Usage

<table>
<thead>
<tr>
<th>Method of Capturing or Preparing Data</th>
<th>Respondents Reporting this Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mapping from other codes</td>
<td>8</td>
</tr>
<tr>
<td>Natural Language Processing</td>
<td>4</td>
</tr>
<tr>
<td>SNOMED CT synonyms for searching</td>
<td>3</td>
</tr>
<tr>
<td>User interface terms mapped to SNOMED CT</td>
<td>3</td>
</tr>
<tr>
<td>Use of subsets or SNOMED CT refsets</td>
<td>1</td>
</tr>
</tbody>
</table>

These figures are based on looking for mention of these techniques in narrative reports. It is likely the actual numbers are higher.
Approaches to SNOMED CT Analytics

1. Subset testing
   – Search records for concepts in a specified predefined subset

2. Subsumption testing
   – Search records for a concept or any of its subtypes

3. Constraint testing
   – Search records for concepts that match criteria including subset membership, subsumption rules and defining relationships with specified values

4. Postcoordinated expression testing
   – Search records for expressions that match criteria including subset membership, subsumption rules and defining relationships with specified values

5. Advanced Description Logic testing
   – Enable the searches to take account of more complex description logic inferences
## Approaches to SNOMED CT Analytics - Usage

<table>
<thead>
<tr>
<th>Type of Use</th>
<th>Respondents Reporting this Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subtype hierarchy or subsumption testing</td>
<td>9</td>
</tr>
<tr>
<td>Subsets or reference set testing</td>
<td>5</td>
</tr>
<tr>
<td>Constraint testing (with defining relationship criteria)</td>
<td>3</td>
</tr>
<tr>
<td>Advanced description logic testing</td>
<td>2</td>
</tr>
</tbody>
</table>

These figures are based on looking for mention of these techniques in narrative reports. It is likely the actual numbers are higher.
Analytic Query Examples from Respondents

- Define cohorts built on criteria such as age, gender, diseases, conditions, medication or other treatments
- User configurable reports of health issues and clinical observations for discharged patients to support treatment decisions and protocol implementation
- Identify subsets of concepts for criteria in decision support modules
- Report age-sex distribution of ER attendances with diagnoses that are subtypes of ‘Traumatic injury’
- Find all patients with disorders in a specific subset of concepts (e.g. cardiac diseases)
- Find all patients with disorders caused by a particular organism (e.g. aspergillus)
Challenges for Clinical Analytics

- **Reliability of patient data**
  - Analytics depends on the quality and consistency of source data
  - User interfaces should make it easy to collect high quality data

- **Concept definition issues**
  - Human understanding may differ from logical meaning

- **Information model structures**
  - Contextual information may be represented in different ways such as:
    - Family history section with code for ‘Asthma’
    - Code for ‘Family history of asthma’
    - Code for ‘asthma’ with type ‘Family history’
  - Meaning depends on how the terminology and information model work together

- **Versioning**
  - A clinical terminology must be maintained to remain relevant but updates may affect analytics results
Recommendations

1. Identify and document your requirements for clinical analytics

Big Data Analytics  Small Data Analytics  Virtuous Circle of Clinical Analytics
Recommendations

2. Consider how features of SNOMED CT can help you address your requirements

- **Concept orientation**: enables consistent representation of clinical ideas
- **Subtype polyhierarchy**: supports analysis across overlapping categories
- **Context representation**: supports consistent interpretation of information
- **Defining relationships**: enable analysis based on different aspects of meaning
- **Description logic definitions**: enable logical inferences
Recommendations

3. Evaluate options for preparing data for analysis in way that meet your requirements

– Is mapping required from other code systems?
– Is use of NLP a feasible option for converting text in records to SNOMED CT?
– Will the results of conversion be of sufficient quality to safely meet the requirements?
Recommendations

4. Evaluate approaches to analysis to identify one that meets your requirements

- Subtype testing is an simple way to meet many requirements
- Subsets allow specific selection criteria to be applied
- Constraints and use of description logic provide a flexible way to meet more complex requirements
Future Directions

• Publication of additional case studies focusing on clinical decision support and other aspects of data analytics
• Guidance on use of SNOMED CT to enhance decision support
• Publication of guidelines on to use of SNOMED CT in other standards (e.g. HL7 FHIR) to support interoperability and enable analytics over data recorded in different record structures
• Greater use of SNOMED CT's features in more vendor products
• Realization of large scale benefits from more effective clinical analytics
A Summary of How Benefits Were Realized for the Value of Health IT

Treatment/Clinical

Enhanced clinical analytics enhances quality of treatment by enabling:

• Research into effective care pathways
• Decision support that assists clinical teams to follow effective pathways

Savings

Improved care delivery is cost-effective as it reduces adverse events and unnecessary repetition of costly interventions
Questions

Contact Details
• David Markwell – dma@snomed.org

Acknowledgement
• Dr Linda Bird – lbi@snomed.org
  ▪ Senior Implementation Specialist, SNOMED International
  ▪ Co-author of this presentation and author of “Analytics with SNOMED CT”

Links
• Analytics with SNOMED CT
  • http://snomed.org/analytics
• Decision Support with SNOMED CT
  • http://snomed.org/cds

Please complete the online session evaluation!