Biomedical Ontology in Practice

Olivier Bodenreider
Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA
Objectives

◆ Learn about biomedical ontologies
  ● History
  ● Design principles, formalisms and tools
  ● What are they?
  ● What are they used for?

◆ Work with biomedical ontologies
  ● Search
  ● Analyze
  ● Extend
  ● Use for data integration
# Agenda

<table>
<thead>
<tr>
<th>Monday, June 9</th>
<th>Introduction to Biomedical Ontologies</th>
<th>Design Principles, Formalisms and Tools for Biomedical Ontologies</th>
<th>Biomedical Ontologies - Content and structure - Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td>Extending Biomedical Ontologies</td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>
References  Bio-ontology courses

• Barry Smith, U. Buffalo / NCBO
  • http://ontology.buffalo.edu/smith/Ontology_Course.html

• Stefan Schulz, U. Freiburg, Germany / KR-MED
  2008 tutorial
  • http://www.kr-med.org/2008/index.html
## Agenda

<table>
<thead>
<tr>
<th>Monday, June 9</th>
<th>Introduction to Biomedical Ontologies</th>
<th>Design Principles, Formalisms and Tools for Biomedical Ontologies</th>
<th>Biomedical Ontologies - Content and structure - Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td>Extending Biomedical Ontologies</td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>
Introduction to Biomedical Ontologies

June 9, 2008 – Session #1

Olivier Bodenreider
Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Outline

- Historical perspective
- Introduction to biomedical terminologies through an example
- Biomedical terms as names for biomedical classes
- Terminological relations as a surrogate for ontological relations
Historical perspective
Why biomedical terminologies?

- To support a theory of diseases
- To classify diseases
- To support epidemiology
- To index and retrieve information
- To serve as a reference
To support a theory of diseases

- Hippocrates
  - Dismisses superstition
  - Four humors
    - Blood
    - Phlegm
    - Yellow bile
    - Black bile

- Thomas Sydenham (1624-1689)
  - Medical observations on the history and cure of acute diseases (1676)
To classify diseases (and plants)

◆ Carolus Linnaeus (1707-1778)
  ● *Genera Plantarum* (1737)
  ● *Genera Morborum* (1763)

◆ François Boissier de La Croix
  a.k.a. F. B. de Sauvages (1706-1767)
  ● *Methodus Foliorum* (1751)
  ● *Nosologia Methodica* (1763/68)

◆ William Cullen (1710-1790)
  ● *Synopsis Nosologiae Methodicae* (1785)
From plants...
... to diseases

◆ Four categories (W. Cullen)
  ● Fevers
  ● Nervous disorders
  ● Cachexias
  ● Local diseases

“The distinction of the genera of diseases, the distinction of the species of each, and often even that of the varieties, I hold to be a necessary foundation of every plan of physic, whether dogmatical or empirical.”
– William Cullen, Edinburgh, 1785
Synopsis Nosologia Methodicae

(Cited by Chris Chute)
To support epidemiology

◆ John Graunt (1620-1674)
  ● Analyzes the vital statistics of the citizens of London

◆ William Farr (1807-1883)
  ● Medical statistician
  ● Improves Cullen’s classification
  ● Contributes to creating ICD

◆ Jacques Berthillon (1851-1922)
  ● Chief of the statistical services (Paris)
  ● Classification of causes of death (161 rubrics)
London Bills of Mortality

A generaill Bill for this present year, ending the 15 of December 1665, according to the Report made to the KINGS most Excellent Majesty, By the Company of Parish-Clerks of London, etc.

A Boutine and Stillborne 817 Executed 11 Pale 16.58
Aged 1545. Flux and Small Pox 653. Plague 685.88
Apoxtic and Fever 578 Found dead in Bridges, fields &c. 558. Plague 5.8
Bredne 54 Fraserd 25 Proctor 6
Brest 32. Gout and Scabs 31.
Bleeding 26 Liver 15.7
Bloody Flux, Scothing & Flux 183 Grippin in the Guts 1238 Kiling of the Lights 15.7
Bums and Scalded 8 Hang'd & made away themselves 12.8
Cancer 12. Gangrene and Filsal 16
Canker, and Thrus 13. Impomous 21.8
Chilled 165. Kill'd by overall accidents 16 Lumber 1.6
Christmes and Infants 125. Kings Evil 14
Cold and Cough 63. Leprose 14 Scoping in the Neck 3.8
Collick and Winde 13. Lethary 1.7
Confliction and Tiffick 48 Flux. Liver-crown 25. Stone and Stangany 3.8
Convulsion and Morter 109. Measles 1.6
Dilfreted 17. Measles 1.6
Doubt and Tumpany 1478. Mothered and Sick 9. Vomiting 1.6
Drowned 13.7 Overd & Staved 3

Males 5114
Females 4853
All 9957

Increase in the Burials in the 15 of Parishes and in the Pest-House this year 715.88.
Increase of the Plague in the 13 of Parishes and at the Pest-House this year 885.98.
Limitations of existing classifications

“The advantages of a uniform statistical nomenclature, however imperfect, are so obvious, that it is surprising no attention has been paid to its enforcement in Bills of Mortality. Each disease has, in many instances, been denoted by three or four terms, and each term has been applied to as many different diseases: vague, inconvenient names have been employed, or complications have been registered instead of primary diseases. The nomenclature is of as much importance in this department of inquiry as weights and measures in the physical sciences, and should be settled without delay.”

– William Farr

First annual report.
To index and retrieve information

- Biomedical literature
  - MEDLINE (15M citations from 4600 journals)
  - Manually indexed
  - Medical Subject Headings (MeSH)
- Genome
  - Model organism databases (Fly, Mouse, Yeast, …)
  - Manually / semi-automatically curated
  - Gene Ontology
Black bile and psychomotor retardation: shades of melancholia in Dante's Inferno.

Widmer DA.

Memorial Sloan-Kettering Cancer Center, New York, NY 10017, USA widmerd@mskcc.org

The history of melancholy depression is rich with images of movement retardation and mental dysfunction. The recent restoration of psychomotor symptoms to the diagnostic terminology of affective disorder is not novel to the students of medieval melancholia. The move back to the biology of this psychomotor dysfunction with the technical advances in brain imaging in recent years only echoes centuries-old writings on the centrality of movement changes in the depressive condition. The Inferno, the first cantica of Dante Alighieri's Commedia, has a wonderful abundance of allusions to the importance of psychomotor symptoms in describing the depressed individual. Slowed steps, garbled speech, frozen tears, these and many other images keep the physical manifestations of psychomotor suffering in the forefront of the reader's mind. Considering Medieval and Renaissance writings on melancholy suffering, it is fitting that Dante shows a bodily illness reflected in the hellish torments visited on the damned. From the souls of the sullen to those of the violent, the panorama of psychomotor symptoms plays a prominent role in the poem as well as in the medical and literary prose of succeeding centuries.

MeSH Terms:
- Depressive Disorder/history*
- History of Medicine, Medieval
- Human
- Italy
- Literature, Medieval/history*
- Medicine in Literature*
- Poetry/history*
- Psychomotor Disorders/history*
Mouse Genome Database and GO

Entrez Gene

1: Nf2 neurofibromatosis 2 [Mus musculus]
GeneID: 18016  Locus tag: MGL97307

General gene information
GeneOntology
Function
- cytoskeletal protein binding
- protein binding
- structural molecule activity
Process
- intercellular junction assembly and/or maintenance
- negative regulation of cell cycle
- negative regulation of protein kinase activity
- regulation of cell proliferation
Component
- adherens junction
- cytoplasm
- cytoskeleton
- membrane

Evidence
- IEA
- IPI  PubMed
- IDA  PubMed
- IMP  PubMed
- IMP  PubMed
- IMP  PubMed
- IMP  PubMed
- IEA
- IEA
- IEA
- IEA
To serve as a reference

- Reference terminology/ontology
  - Universally needed
  - Developed independently of any purposes
  - Reusable by many applications

- Examples
  - VA National Drug File (NDF)
  - Foundational Model of Anatomy (FMA)
  - SNOMED CT
Anatomy in Biomedicine

- Physiology
- Clinical medicine
- Biomedical literature
- Biomedical research
Administrative terminologies

- Coding patient records
  - International Classification of Primary Care (ICPC)
  - SNOMED
  - Read Codes

- Reporting claims to health insurance companies
  - International Classification of Diseases (ICD-9 CM)
  - Healthcare Common Procedure Coding System (HCPCS)
History of Medical Ontologies

Synopsis
Nosologiae
Methodicae

<table>
<thead>
<tr>
<th>Year</th>
<th>Ontology</th>
</tr>
</thead>
<tbody>
<tr>
<td>1603</td>
<td>Synopsis</td>
</tr>
<tr>
<td>1700</td>
<td>MeSH</td>
</tr>
<tr>
<td>1785</td>
<td>ICD</td>
</tr>
<tr>
<td>1855</td>
<td>ICD9</td>
</tr>
<tr>
<td>1900</td>
<td>SNOP</td>
</tr>
<tr>
<td>1975</td>
<td>NLM</td>
</tr>
</tbody>
</table>

(courtesy of J. Rogers)
Biomedical ontology in PubMed

Number of articles in PubMed/MEDLINE on Ontology vs. Controlled vocabulary

- Ontology or ontologies
- Both
- Controlled vocabulary [excluding DSM]

(*) As of 2008/02/20 (Partial coverage for 2007, due to a slight lag in the indexing process)

[Bodenreider, YBMI 2008]
Biomedical ontologies in PubMed

Proportion of citations in PubMed/MEDLINE by ontology

- GO
- NCI Thesaurus
- FMA
- UMLS
- SNOMED
- MeSH
- LOINC

[Bodenreider, YBMI 2008]
Introduction to biomedical terminologies through an example
Guy’s Hospital, London
Thomas Addison (1795-1860)
Addison’s disease

- Addison's disease is a rare endocrine disorder
- Addison's disease occurs when the adrenal glands do not produce enough of the hormone cortisol
- For this reason, the disease is sometimes called chronic adrenal insufficiency, or hypocortisolism
Adrenal insufficiency  Clinical variants

◆ Primary / Secondary
  ● Primary: lesion of the adrenal glands themselves
  ● Secondary: inadequate secretion of ACTH by the pituitary gland

◆ Acute / Chronic

◆ Isolated / Polyendocrine deficiency syndrome
Addison’s disease: Symptoms

- Fatigue
- Weakness
- Low blood pressure
- Pigmentation of the skin (exposed and non-exposed parts of the body)
- ...

Lister Hill National Center for Biomedical Communications
AD in medical vocabularies

◆ Synonyms: different terms
  ● Addisonian syndrome
  ● Bronzed disease
  ● Addison melanoderma
  ● Asthenia pigmentosa
  ● Primary adrenal deficiency
  ● Primary adrenal insufficiency
  ● Primary adrenocortical insufficiency
  ● Chronic adrenocortical insufficiency

◆ Contexts: different hierarchies

Lister Hill National Center for Biomedical Communications
### CHAPTER 4
Endocrine, nutritional and metabolic diseases (E00-E90)

#### Disorders of other endocrine glands (E20-E35)

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>E27</td>
<td>Other disorders of adrenal gland</td>
</tr>
<tr>
<td>E27.0</td>
<td>Other adrenocortical overactivity</td>
</tr>
<tr>
<td></td>
<td>Overproduction of ACTH, not associated with Cushing’s disease</td>
</tr>
<tr>
<td></td>
<td>Premature adrenarche</td>
</tr>
<tr>
<td></td>
<td>Excludes 1: Cushing’s syndrome (E24.-)</td>
</tr>
<tr>
<td>E27.1</td>
<td>Primary adrenocortical insufficiency</td>
</tr>
<tr>
<td></td>
<td>Addison’s disease</td>
</tr>
<tr>
<td></td>
<td>Adrenocortical insufficiency NOS</td>
</tr>
<tr>
<td></td>
<td>Autoimmune adrenalitis</td>
</tr>
<tr>
<td></td>
<td>Excludes 1: Addison only phenotype adrenoleukodystrophy (E71.428)</td>
</tr>
<tr>
<td></td>
<td>Amyloidosis (E85)</td>
</tr>
<tr>
<td></td>
<td>Tuberculous Addison’s disease (A18.7)</td>
</tr>
<tr>
<td></td>
<td>Waterhouse-Friderichsen syndrome (A39.1)</td>
</tr>
<tr>
<td>E27.2</td>
<td>Addisonian crisis</td>
</tr>
<tr>
<td></td>
<td>Adrenal crisis</td>
</tr>
<tr>
<td></td>
<td>Adrenocortical crisis</td>
</tr>
<tr>
<td>E27.3</td>
<td>Drug-induced adrenocortical insufficiency</td>
</tr>
<tr>
<td></td>
<td>Code first (T36-T50) to identify drug</td>
</tr>
<tr>
<td>E27.4</td>
<td>Other and unspecified adrenocortical insufficiency</td>
</tr>
</tbody>
</table>

Draft ICD-10-CM Tabular Page 180 June 2003
Medical Subject Headings

MeSH Tree Structures

Endocrine Diseases [C:9]
  Adrenal Gland Diseases [C19.053]
    Adrenal Gland Hypofunction [C19.053.264]
      ▶ Addison’s Disease [C19.053.264.263]
      Adrenoleukodystrophy [C19.053.264.270]
      Hypoaldosteronism [C19.053.264.480]

Immunologic Diseases [C:20]
  Autoimmune Diseases [C20.111]
    ▶ Addison’s Disease [C20.111.163]
      Anemia, Hemolytic, Autoimmune [C20.111.175]
      Anti-Glomerular Basement Membrane Disease [C20.111.190]
      Antiphospholipid Syndrome [C20.111.197]
      Arthritis, Rheumatoid [C20.111.199] +
      Autoimmune Diseases of the Nervous System [C20.111.258] +
Addison's disease - Definition

Concept Status: Current

Descriptions
-  Addison's disease (disorder)
-  Addison's disease
-  Enfermedad de Addison
-  Enfermedad de Addison (trastorno)

Definition: Primitive
-  is a
  -  adrenal cortical hypofunction
  -  finding site
  -  adrenal cortex structure

Qualifiers
-  severity
  -  severities
-  episodicity
  -  episodicities
-  clinical course
  -  courses

Codes
-  Original SnomedId : DB-70620
-  Read Code (Ctv3Id) : C1541
Biomedical terms as names for biomedical classes
Terms reflecting valid classes

- Pulmonary anthrax
- BRCA1 protein
- Coronary artery
- Coronary artery bypass
- ...

- Non-insulin dependent diabetes mellitus
- Non-Hodgkin lymphoma
- Non-steroidal anti-inflammatory drugs
- Non-opioid analgesics
- Non-invasive medical procedure
Issues

- Multiple terms for a class
- Multiple classes for a term
- Presence of non-ontological features in terms
- Composite terms
Multiple terms for a class

◆ Synonymy
-
  - Left coronary artery
  - LCA
  - Arteria coronaria sinistra
-
  - Addison’s disease
  - Primary adrenocortical insufficiency

◆ “Clinical synonymy” (vs. identity)
-
  - Abdominal swelling
  - Swollen abdomen
-
  - Posttransfusion hepatitis
  - Posttransfusion viral hepatitis
-
  - Addison’s disease
  - Primary adrenocortical insufficiency
  - vs. Waterhouse-Friderichsen Syndrome
Multiple classes for a term

- **Polysemy**
  - Cold
    - Cold
    - Common cold
    - Cold temperature
    - COLD
    - Chronic Obstructive Airway Disease

- **Truncated terms**
  - Calcium
    - Ca++
    - Coagulation factor IV
    - Calcium
    - Calcium measurement
Non-ontological features in terms

- Epistemological features
  - Gallbladder calculus without mention of cholecystitis
  - Diarrhea of presumed infectious origin
  - Replacement of unspecified heart valve
  - ...

Lister Hill National Center for Biomedical Communications
Ontology vs. Epistemology

- **Ontology**
  - Invariants in reality
    - Classes (universals)
    - Relations between them
  - Theory of reality

- **Epistemology**
  - Knowledge about such entities
  - Perception of reality

Bone metastasis
diagnosed by CT scan

Bone metastasis
diagnosed by Tc99m bone scintiscan
Composite terms

- **Sentence-like terms**
  - Several classes and their relations
  - May contain epistemological features

- Tuberculosis of adrenal glands, tubercle bacilli not found (in sputum) by microscopy, but found by bacterial culture
More composite terms

- Nontraffic accident involving being accidentally pushed from motor vehicle, except off-road motor vehicle, while in motion, not on public highway, driver of motor vehicle injured

- Determine whether the elder patient and caretaker have a functional social support network to assist the patient in performing activities of daily living and in obtaining health care, transportation, therapy, medications, community resource information, financial advice, and assistance with personal problems

- Telephone call by a physician to patient or for consultation or medical management or for coordinating medical management with other health care professionals (eg, nurses, therapists, social workers, nutritionists, physicians, pharmacists); complex or lengthy (eg, lengthy counseling session with anxious or distraught patient, detailed or prolonged discussion with family members regarding seriously ill patient, lengthy communication necessary to coordinate complex services of several different health professionals working on different
Terminological relations as a surrogate for ontological relations
Issues

◆ Lack of explicit classificatory principle
◆ Underspecification of the relations
◆ Thesaurus relations
◆ Limited depth in hierarchies “by design”
Explicit classificatory principle

Foundational Model of Anatomy

Anatomical entity

Physical anatomical entity

Material physical anatomical entity

Anatomical structure

Body substance

Non-material physical anatomical entity

Anat. space

Anat. surface

Anat. line

Anat. point

3D

2D

1D

0D

Spatial dimension

Mass

Inherent 3D shape
No explicit classificatory principle

3. Diseases [C]
   - Bacterial Infections and Mycoses [C01]
   - Virus Diseases [C02]
   - Parasitic Diseases [C03]
   - Neoplasms [C04]
   - Musculoskeletal Diseases [C05]
   - Digestive System Diseases [C06]
   - Stomatognathic Diseases [C07]
   - Respiratory Tract Diseases [C08]
   - Otorhinolaryngologic Diseases [C09]
   - Nervous System Diseases [C10]
   - Eye Diseases [C11]
   - Urologic and Male Genital Diseases [C12]
   - Female Genital Diseases and Pregnancy Complications [C13]
   - Cardiovascular Diseases [C14]
   - Hemic and Lymphatic Diseases [C15]
   - Neonatal Diseases and Abnormalities [C16]
   - Skin and Connective Tissue Diseases [C17]
   - Nutritional and Metabolic Diseases [C18]
   - Endocrine Diseases [C19]
   - Immunologic Diseases [C20]
   - Disorders of Environmental Origin [C21]
   - Animal Diseases [C22]
   - Pathological Conditions, Signs and Symptoms [C23]
1. Certain infectious and parasitic diseases
2. Neoplasms
3. Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
4. Endocrine, nutritional, and metabolic diseases
5. Mental and behavioral disorders
6. Diseases of nervous system
7. Diseases of the eye and adnexa
8. Diseases of the ear and mastoid process
9. Diseases of circulatory system
10. Diseases of respiratory system
11. Diseases of digestive system
12. Diseases of the skin and subcutaneous tissue
13. Diseases of the musculoskeletal system and connective tissue
14. Diseases of the genitourinary system
15. Pregnancy, childbirth, and the puerperium
16. Certain conditions originating in the newborn (perinatal) period
17. Congenital malformations, deformations and chromosomal abnormalities
18. Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
19. Injury, poisoning and certain other consequences of external causes
20. External causes of morbidity
21. Factors influencing health status and contact with health service
- Attribute
- Body structure
- Clinical finding
- Context-dependent categories
- Environments and geographical locations
- Events
- Observable entity
- Organism
- Pharmaceutical / biologic product
- Physical force
- Physical object
- Procedure
- Qualifier value
- Social context
- Special concept
- Specimen
- Staging and scales
- Substance
Fully specified relations

Viral meningitis in SNOMED CT

- Fully defined by ...
  - Is a
    - Dviral infections of the central nervous system
    - Dinf ective meningitis
  - Causative agent
    - Dvirus
  - Group
    - Associated morphology
      - Dinflammation
    - Finding site
      - Dmeninges structure
Underspecification of the relations

Diseases

CNS diseases
CNS viral diseases

CNS infections

Meningitis
Viral meningitis

Virus diseases

parent
child

isa ?
Thesaurus relations

- Addison’s disease
  - Due to auto-immunity in 80% of the cases
  - Other causes include tuberculosis

Relations used to create hierarchical structures vs. hierarchical relations
Endocrine Diseases [C19]

Adrenal Gland Diseases [C19.053]

Adrenal Gland Hypofunction [C19.053.264]

▶ Addison's Disease [C19.053.264.263]

Adrenoleukodystrophy [C19.053.264.270]

Hypoaldosteronism [C19.053.264.480]

Immunologic Diseases [C20]

Autoimmune Diseases [C20.111]

▶ Addison's Disease [C20.111.163]

Anemia, Hemolytic, Autoimmune [C20.111.175]

Anti-Glomerular Basement Membrane Disease [C20.111.190]

Antiphospholipid Syndrome [C20.111.197]

Arthritis, Rheumatoid [C20.111.199]

Hierarchy

- adrenal cortical hypofunction
  - Addison's disease
    - Addison's disease due to autoimmunity
    - Addison's disease with adrenoleukodystrophy
    - polyglandular autoimmune syndrome, type 1
    - tuberculous Addison's disease
Accidents in MeSH

Environment and Public Health [G03]
Public Health [G03.850]

Accidents [G03.850.110]

Accident Prevention [G03.850.110.060] +
Accidental Falls [G03.850.110.085]
Accidents, Aviation [G03.850.110.185]
Accidents, Home [G03.850.110.205]
Accidents, Occupational [G03.850.110.250] +
Accidents, Radiation [G03.850.110.285]
Accidents, Traffic [G03.850.110.320]
Drowning [G03.850.110.500] +
Limited depth in hierarchies “by design”

◆ Term identifier (code) used to record the position in the hierarchy
  ● Limited number of digits available
  ● May hide part of the structure
◆ Terminologies: ICD, SNOMED, …

E84 Cystic fibrosis
  Includes: mucoviscidosis
  E84.0 Cystic fibrosis with pulmonary manifestations
    Use additional code to identify any infectious organism present, such as:
    Pseudomonas (B96.5)
  E84.1 Meconium ileus in cystic fibrosis
    Excludes1: meconium ileus not due to Cystic fibrosis (P75)
  E84.2 Cystic fibrosis with gastrointestinal manifestations
    Excludes2: meconium ileus in cystic fibrosis (E84.1)
  E84.8 Cystic fibrosis with other manifestations
Cystic fibrosis in ICD

E84 Cystic fibrosis
Includes: mucoviscidosis
E84.0 Cystic fibrosis with pulmonary manifestations
   Use additional code to identify any infectious organism present, such as:
   Pseudomonas (B96.5)
E84.1 Meconium ileus in cystic fibrosis
   Excludes1: meconium ileus not due to Cystic fibrosis (P75)
E84.2 Cystic fibrosis with gastrointestinal manifestations
   Excludes2: meconium ileus in cystic fibrosis (E84.1)
E84.8 Cystic fibrosis with other manifestations
Conclusions
Conclusions 😞

- Biomedical terms
  - reflect some aspects of biomedical reality
    - Although the primary concern of terminology is naming, not reflecting reality
  - often convey additional features (e.g., epistemology)
- Biomedical terminology tends to offset part of the complexity
  - but often reflects utility
Conclusions 😊

- Biomedical terminologies can help populate biomedical ontologies

- Resources needed
  - Linguistic analysis of terms
  - Statistical analysis of terms in a corpus / annotation database (dependence relations)
  - Manual curation
# Agenda

<table>
<thead>
<tr>
<th>Monday, June 9</th>
<th>Introduction to Biomedical Ontologies</th>
<th>Design Principles, Formalisms and Tools for Biomedical Ontologies</th>
<th>Biomedical Ontologies - Content and structure - Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td>Extending Biomedical Ontologies</td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>

Lister Hill National Center for Biomedical Communications
Design Principles, Formalisms and Tools for Biomedical Ontologies

Olivier Bodenreider
Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Overview

◆ Definitions
  ● Ontologies vs. other artifacts
  ● Kinds of ontologies

◆ Some principles of formal ontology
  ● Top-level categories
  ● Categories of relationships

◆ Formalisms and tools
Definitions
Introduction

Symbol

Concept

Object

Ogden-Richards
Definitions

◆ The *What* question
  ● Objects in the world
    ■ With their properties
    ■ With their relations to other objects
  ● Also: events, processes, and states

◆ Explicit specification of a conceptualization
  ● Support software applications

◆ Domain ontology reflects
  ● Underlying reality
  ● Theory of the domain
Examples of use

- Natural language processing
- Access to heterogeneous sources of information (e.g., Semantic Web)
- Systems engineering
  
- Interoperability
Ontology vs. other artifacts

- **Ontology**
  - Defining types of things and their relations

- **Terminology**
  - Naming things in a domain

- **Thesaurus**
  - Organizing things for a given purpose

- **Classification**
  - Placing things into (arbitrary) classes

- **Knowledge bases**
  - Assertional knowledge

[Smith, KR-MED 2006]
[Chute, JAMIA 2000]
(Controlled) Terminology

- **Objective:** naming things
- **Example:** Current Procedural Terminology (CPT)
- **Shared understanding**
  - Agreement on what terms to use
  - Utility-driven (arbitrary)

Telephone call by a physician to patient or for consultation or medical management or for coordinating medical management with other health care professionals (e.g., nurses, therapists, social workers, nutritionists, physicians, pharmacists); complex or lengthy (e.g., lengthy counseling session with anxious or distraught patient, detailed or prolonged discussion with family members regarding seriously ill patient, lengthy communication necessary to coordinate complex services of several different health professionals working on different
Thesaurus

◆ Objective: organize things for a purpose
  • e.g., information retrieval
    ■ Organization by relatedness

◆ Example: Medical Subject Headings (MeSH)
  • Indexing/retrieval of biomedical articles

◆ Relations used in hierarchies
  vs. hierarchical relations
Thesaurus vs. ontology

Autoimmune Diseases

is generally a

Addison’s disease

Tuberculous Addison’s disease

Addison’s disease due to autoimmunity
Classification

◆ Objective: placing things into classes
◆ Characteristics
  • Single inheritance (tree)
  • Idiosyncratic inclusion/exclusion criteria

Insulin-dependent diabetes mellitus
[See before E10 for subdivisions.]
Includes: diabetes (mellitus):
  • brittle
  • juvenile-onset
  • ketosis-prone
  • type I
Excludes: diabetes mellitus (in):
  • malnutrition-related (E12.-)
  • neonatal (P70.2)
  • pregnancy, childbirth and the puerperium (O24.-)
glycosuria:
  • NOS (R81.)
  • renal (E74.8)
impaired glucose tolerance (R73.0)
postsurgical hypoinsulinaemia (E89.1)
Classification

◆ Characteristics (continued)
  - Everything must be classified, including
    - When there is no specific slot (NEC)
    - When there is insufficient information (NOS)

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>E84</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td></td>
<td><em>Includes</em>: mucoviscidosis</td>
</tr>
<tr>
<td>E84.0</td>
<td>Cystic fibrosis with pulmonary manifestations</td>
</tr>
<tr>
<td>E84.1</td>
<td>Cystic fibrosis with intestinal manifestations</td>
</tr>
<tr>
<td></td>
<td>Meconium ileus+ (P75*)</td>
</tr>
<tr>
<td></td>
<td><em>Excludes</em>: meconium obstruction in cases where</td>
</tr>
<tr>
<td></td>
<td>cystic fibrosis is known not to be present (P76.0)</td>
</tr>
<tr>
<td>E84.8</td>
<td>Cystic fibrosis with other manifestations</td>
</tr>
<tr>
<td>E84.9</td>
<td>Cystic fibrosis, unspecified</td>
</tr>
</tbody>
</table>
Knowledge Bases

◆ Objective: represent knowledge needed for a given application
◆ Example: drug knowledge bases
◆ Assertional knowledge
  ● Vs. definitional knowledge in ontologies
  ● Often probabilistic
◆ Examples of assertions
  ● Indications of a drug
  ● Signs and symptoms of a disease
Fuzzy borders

- Some ontologies also collect names
  - FMA
- Some terminologies also provide formal definitions
  - SNOMED CT
- Some terminologies/ontologies include both definitional and assertional knowledge
  - SNOMED CT
Types of resources

- **Lexical resources**
  - Collections of lexical items
  - Additional information
    - Part of speech
    - Spelling variants
  - Useful for entity recognition
  - UMLS SPECIALIST Lexicon, WordNet

- **Ontological resources**
  - Collections of
    - kinds of entities
      - substances, qualities, processes
    - relations among them
  - Useful for relation extraction
  - UMLS Semantic Network, BioTop

- **Terminological resources**
  - Collections lexical items + identifiers
    - Useful for entity resolution
    - UMLS Metathesaurus
The Knowledge Semantics Continuum

Ontology Dimensions based on McGuinness and Finin
Kinds of ontologies

- Upper Level Ontology
- General Ontology
- Domain Ontology

Application ontologies
Ontology-related issues

- Creation
- Merging
- Alignment
- Validation
Formal Ontological Principles
Formal ontological distinctions

- Universal vs. individual
- Continuant vs. occurrent
- Independent vs. dependent
Universal vs. Individual

- **Universal** = *category*
- **Synonyms**
  - Kind, Type, (Class)
- **Examples**
  - eyeball
  - blood pressure
  - conference

- **Individual** = *instance*
- **Synonyms**
  - Particular, Token
- **Examples**
  - my right eyeball
  - my blood pressure (132/79)
  - AMIA Annual Symposium 2003

*instantiation*
Continuant vs. Occurrent

- Continuant = *Continues to exist through time*
  - Synonyms
    - Substance
  - Examples
    - tennis racquet
    - mitochondrion
    - insulin production

- Occurrent = *Unfolds through time*
  - Synonyms
    - Process
  - Examples
    - tennis tournament
    - metabolism
    - producing insulin
Independent vs. Dependent

- **Independent** = Can exist without support from other entities
- **Examples**
  - virus
  - molecule
  - plant

- **Dependent** = Require support from other entities for its existence
- **Examples**
  - viral infection
  - DNA binding
  - food
Formal ontology  Upper level

Universals  (classes)

- Continuant
  - Independent continuant
  - Dependent continuant

Occurent

Particulars  (instances)

Lister Hill National Center for Biomedical Communications
Formal ontological distinctions

- Basic distinctions in many top-level ontologies
  - Generic: BFO, DOLCE
  - Biomedical: BioTop, UMLS Semantic Network
- Condition the relations between various types of entities
  - Relations
    - Between instances (e.g., part_of [at time])
    - Between classes (e.g., isa, part_of [attemporal])
    - Between one instance and one class (instance_of)

[Smith, Genome Biology 2005]
Formal ontology in practice

- Provides foundational classes and relations
  - Upper level ontologies
  - Relation ontology
- Provides a framework for analyzing entities and relations
Examples
General ontologies

◆ OpenCyc
  ● General ontology
  ● Cycorp, Inc (D. Lenat & al.)
  ● Over 1M hand-coded assertions
  ● http://www.opencyc.org

◆ WordNet
  ● Electronic lexical database
  ● Princeton University (G. Miller & al.)
  ● Over 100,000 synsets
  ● http://wordnet.princeton.edu/
Top level in OpenCyc

- Thing
  - Intangible
    - Mathematical or computational thing
    - Set or collection
      - Collection
  - Intangible individual
  - Individual
  - PartiallyTangible
    - Tangible thing

Diagram: #genls, #isa
Top level in WordNet

Abstraction
Activity
Entity
Event
Group
Location
Natural phenomenon
Possession
Psychological feature
Shape
State
Generalised Architecture for Languages, Encyclopaedias, and Nomenclatures in Medicine

European Union project (A. Rector & al.)

Over 25,000 concepts (primitives)

http://www.opengalen.org
Top level in GALEN
UMLS Semantic Network

- Definitional knowledge in the biomedical domain
- NLM (A. McCray & al.)
- Content
  - 135 semantic types
  - 54 types of relationship
  - 6700 semantic relations
Top level in the Semantic Network

- Root
  - Entity
    - Conceptual Entity
    - Physical object
  - Event
    - Activity
    - Phenomenon or Process
Differences between ontologies

Examples
Granularity, plesionymy

UMLS

Epilepsy, Generalized
Seizure Disorder, Generalized
[...]

Epilepsy, Grand Mal
Tonic-Clonic Epilepsy
Seizure Disorder, Tonic Clonic
[...]

WordNet

generalized epilepsy
grand mal epilepsy
Differing categorization

UMLS

Natural Phenomenon or Process

Biologic Function

Pathologic Function

Health disorder

Disease or Syndrome

Dental Caries
Dental cavity, NOS
Tooth caries
Dental Decay

WordNet

phenomenon

process

natural process

decay

cavity caries
dental caries
tooth decay

Dental Caries
Formalisms and Tools
Ontology and Formalism

- Frames
- Description logics
  - OWL DL
- First-order logic

- OBO Format
  - Conversion to OWL DL
Tools for ontology developers

◆ Protégé
  - Publicly available
  - Frames and DL
  - Classifier
  - Supports many file formats (import/export)
  - Large community of users
  - Well supported
  - http://protege.stanford.edu/

◆ OBO-Edit
  - Specific to the biomedical/OBO community
  - Simpler than Protégé (“tool for biologists”)
  - http://oboedit.org/

Lister Hill National Center for Biomedical Communications
## Agenda

<table>
<thead>
<tr>
<th>Monday, June 9</th>
<th>Introduction to Biomedical Ontologies</th>
<th>Design Principles, Formalisms and Tools for Biomedical Ontologies</th>
<th>Biomedical Ontologies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Content and structure</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Function</td>
</tr>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td>Extending Biomedical Ontologies</td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>
“High-Impact” Biomedical Ontologies

A Structural Perspective

Olivier Bodenreider

Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Overview

◆ Structural perspective
  ● What are they (vs. what are they for)?

◆ “High-impact” biomedical ontologies
  ● International Classification of Diseases (ICD)
  ● Logical Observation Identifiers, Names and Codes (LOINC)
  ● SNOMED Clinical Terms
  ● Foundational Model of Anatomy
  ● Gene Ontology
  ● RxNorm
  ● Medical Subject Headings (MeSH)
  ● NCI Thesaurus
  ● Unified Medical Language System (UMLS)

[J. Cimino, YBMI 2006]
ICD Characteristics (1)

- Current version: ICD-10
- Type: Classification
- Domain: Disorders
- Developer: World Health Organization (WHO)
- Funding: WHO
- Availability
  - Publicly available: No
  - Repositories: UMLS  [ICD9-CM in NCBO BioPortal]
- URL: http://www.who.int/classifications/icd/en/
ICD Characteristics (2)

◆ Number of
  ● Concepts: 12,318
  ● Terms: 1 per concept (tabular)

◆ Major organizing principles:
  ● Tree (single inheritance hierarchy)
  ● No explicit classification criteria
    ■ Idiosyncratic inclusion/exclusion mechanism
  ● .8 slots for Not elsewhere classified (NEC)
  ● .9 slots for Not otherwise specified (NOS)

◆ Formalism: Proprietary format
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Blocks</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>A00-B99</td>
<td>Certain infectious and parasitic diseases</td>
</tr>
<tr>
<td>II</td>
<td>C00-D48</td>
<td>Neoplasms</td>
</tr>
<tr>
<td>III</td>
<td>D50-D89</td>
<td>Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism</td>
</tr>
<tr>
<td>IV</td>
<td>E00-E90</td>
<td>Endocrine, nutritional and metabolic diseases</td>
</tr>
<tr>
<td>V</td>
<td>F00-F99</td>
<td>Mental and behavioural disorders</td>
</tr>
<tr>
<td>VI</td>
<td>G00-G99</td>
<td>Diseases of the nervous system</td>
</tr>
<tr>
<td>VII</td>
<td>H00-H59</td>
<td>Diseases of the eye and adnexa</td>
</tr>
<tr>
<td>VIII</td>
<td>H60-H95</td>
<td>Diseases of the ear and mastoid process</td>
</tr>
<tr>
<td>IX</td>
<td>I00-I99</td>
<td>Diseases of the circulatory system</td>
</tr>
<tr>
<td>X</td>
<td>J00-J99</td>
<td>Diseases of the respiratory system</td>
</tr>
<tr>
<td>XI</td>
<td>K00-K93</td>
<td>Diseases of the digestive system</td>
</tr>
<tr>
<td>XII</td>
<td>L00-L99</td>
<td>Diseases of the skin and subcutaneous tissue</td>
</tr>
<tr>
<td>XIII</td>
<td>M00-M99</td>
<td>Diseases of the musculoskeletal system and connective tissue</td>
</tr>
<tr>
<td>XIV</td>
<td>N00-N99</td>
<td>Diseases of the genitourinary system</td>
</tr>
<tr>
<td>XV</td>
<td>O00-O99</td>
<td>Pregnancy, childbirth and the puerperium</td>
</tr>
<tr>
<td>XVI</td>
<td>P00-P96</td>
<td>Certain conditions originating in the perinatal period</td>
</tr>
<tr>
<td>XVII</td>
<td>Q00-Q99</td>
<td>Congenital malformations, deformations and chromosomal abnormalities</td>
</tr>
<tr>
<td>XVIII</td>
<td>R00-R99</td>
<td>Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified</td>
</tr>
<tr>
<td>XIX</td>
<td>S00-T98</td>
<td>Injury, poisoning and certain other consequences of external causes</td>
</tr>
<tr>
<td>XX</td>
<td>V01-Y98</td>
<td>External causes of morbidity and mortality</td>
</tr>
<tr>
<td>XXI</td>
<td>Z00-Z99</td>
<td>Factors influencing health status and contact with health services</td>
</tr>
<tr>
<td>XXII</td>
<td>U00-U99</td>
<td>Codes for special purposes</td>
</tr>
</tbody>
</table>
ICD Example

◆ Idiosyncratic inclusion/exclusion criteria

**E10**  
*Insulin-dependent diabetes mellitus*  
[See before E10 for subdivisions]

*Includes:* diabetes (mellitus):
  - brittle
  - juvenile-onset
  - ketosis-prone
  - type I

*Excludes:* diabetes mellitus (in):
  - malnutrition-related (E12.-)
  - neonatal (P70.2)
  - pregnancy, childbirth and the puerperium (O24.-)
  - glycosuria:
    - NOS (R81.)
    - renal (E74.8)
  - impaired glucose tolerance (R73.0)
  - postsurgical hypoinsulinaemia (E89.1)
ICD Example

- Not elsewhere classified (NEC)
- Not otherwise specified (NOS)

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>E84</td>
<td>Cystic fibrosis</td>
</tr>
<tr>
<td>E84.0</td>
<td>Cystic fibrosis with pulmonary manifestations</td>
</tr>
<tr>
<td>E84.1</td>
<td>Cystic fibrosis with intestinal manifestations</td>
</tr>
<tr>
<td>E84.8</td>
<td>Cystic fibrosis with other manifestations</td>
</tr>
<tr>
<td>E84.9</td>
<td>Cystic fibrosis, unspecified</td>
</tr>
</tbody>
</table>

Includes: mucoviscidosis

Excludes: meconium obstruction in cases where cystic fibrosis is known not to be present (P76.0)
Logical Observation Identifiers, Names and Codes (LOINC)
LOINC Characteristics (1)

- Current version: 2.22 (Dec. 2007)
- Type: Controlled terminology*
- Domain: Laboratory and clinical observations
- Developer: Regenstrief Institute
- Funding: NLM

Availability
  - Publicly available: Yes
  - Repositories: UMLS

URL: www.regenstrief.org/loinc/loinc.htm
LOINC Characteristics (2)

- **Number of**
  - Concepts: 50k active codes (2.18)
  - Terms: n/a*

- **Major organizing principles:**
  - No hierarchical structure among the main codes
  - 6 axes
    - Component (analyte [+ challenge] [+ adjustments])
    - Property
    - Timing
    - System
    - Scale
    - [Method]

- **Formalism: “DL-like”**
LOINC Example

◆ Sodium:SCnc:-Pt:Ser/Plas:Qn
[the molar concentration of sodium is measured in the plasma (or serum), with quantitative result]

<table>
<thead>
<tr>
<th>Axis</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
<td>Sodium</td>
</tr>
<tr>
<td>Property</td>
<td>SCnc – Substance Concentration (per volume)</td>
</tr>
<tr>
<td>Timing</td>
<td>Pt – Point in time (Random)</td>
</tr>
<tr>
<td>System</td>
<td>Ser/Plas – Serum or Plasma</td>
</tr>
<tr>
<td>Scale</td>
<td>Qn – Quantitative</td>
</tr>
<tr>
<td>Method</td>
<td>--</td>
</tr>
</tbody>
</table>
SNOMED Clinical Terms
SNOMED CT  Characteristics (1)

◆ Current version: January 31, 2008 (2 annual releases)
◆ Type: Reference terminology / ontology
◆ Domain: Clinical medicine
◆ Developer: IHTSDO
◆ Funding: IHTSDO
◆ Availability
  ● Publicly available: Yes* (in member countries)
  ● Repositories: UMLS
◆ URL: http://www.ihtsdo.org/
SNOMED CT Characteristics (2)

◆ Number of

- Concepts: 311,313 active concepts (Jan. 31, 2008)
- Terms: 794,061 active “descriptions”

◆ Major organizing principles:

- Utility for clinical medicine (e.g., assertional + definitional knowledge)
- Model of meaning (incomplete)
- Rich set of associative relationships
- Small proportion of defined concepts (many primitives)

◆ Formalism: Description logics (KRSS)
SNOMED CT  Top level

<table>
<thead>
<tr>
<th>Hierarchy</th>
<th>Subtype hierarchy</th>
</tr>
</thead>
<tbody>
<tr>
<td>138875005</td>
<td>SNOMED CT Concept</td>
</tr>
<tr>
<td>362981000</td>
<td>qualifier value</td>
</tr>
<tr>
<td>106237007</td>
<td>linkage concept</td>
</tr>
<tr>
<td>370115009</td>
<td>special concept</td>
</tr>
<tr>
<td>48176007</td>
<td>social context</td>
</tr>
<tr>
<td>419891008</td>
<td>record artifact</td>
</tr>
<tr>
<td>363787002</td>
<td>observable entity</td>
</tr>
<tr>
<td>308916002</td>
<td>environment or geographical location</td>
</tr>
<tr>
<td>123038009</td>
<td>specimen</td>
</tr>
<tr>
<td>254291000</td>
<td>staging and scales</td>
</tr>
<tr>
<td>123037004</td>
<td>body structure</td>
</tr>
<tr>
<td>272379006</td>
<td>event</td>
</tr>
<tr>
<td>78621006</td>
<td>physical force</td>
</tr>
<tr>
<td>404684003</td>
<td>clinical finding</td>
</tr>
<tr>
<td>260787004</td>
<td>physical object</td>
</tr>
<tr>
<td>410607006</td>
<td>organism</td>
</tr>
<tr>
<td>71388002</td>
<td>procedure</td>
</tr>
<tr>
<td>373873005</td>
<td>pharmaceutical / biologic product</td>
</tr>
<tr>
<td>243796009</td>
<td>situation with explicit context</td>
</tr>
<tr>
<td>105590001</td>
<td>substance</td>
</tr>
</tbody>
</table>
SNOMED CT Example

Hierarchy | Subtype hierarchy
---|---
27010001 | partial excision of large intestine
8613002 | operation on appendix
80146002 | appendectomy

- C 82730006 | incidental appendectomy
- C 49438003 | appendectomy with drainage
- C 174036004 | emergency appendectomy
- C 174045003 | interval appendectomy
- C 6025007 | laparoscopic appendectomy
- C 235313004 | non-emergency appendectomy
- C 235314005 | inversion appendectomy
- C 1299000 | excision of appendiceal stump

Definition: Fully defined by ...

is a
- D partial excision of large intestine
- D operation on appendix

Group
- method
- D excision - action
- procedure site - Direct
- D appendix structure

Qualifiers
- access
- p surgical access values
- priority
- p priorities

Codes
- Original SnomedId : P1-57450
- Read Code (Ctv3Id) : X20Wz
Foundational Model of Anatomy
FMA Characteristics (1)

- Current version: ? (no fixed release schedule)
- Type: Ontology
- Domain: Anatomy (anatomical structures)
- Developer: U. Washington, Department of Biological Structure
- Funding: NLM (grants and contract) and others
- Availability
  - Publicly available: Yes
  - Repositories: [UMLS] / OBO / NCBO BioPortal
- URL: http://fma.biostr.washington.edu/
FMA Characteristics (2)

- Number of
  - Concepts: ~72k
  - Terms: ~1.5 / concept

- Major organizing principles:
  - Explicit classificatory criteria
  - Distinct *isa* and *part_of* hierarchies
  - Additional spatial relations (e.g., adjacency)
  - Multiple levels of granularity (organism to sub-cellular)

- Formalism: Frames (Protégé)
  - Conversion to OWL Full and OWL DL available
FMA Top level

(Courtesy of C. Rosse)
FMA Example

(Courtesy of C. Rosse)
Gene Ontology
Gene Ontology Characteristics (1)

- Current version: n/a (daily/monthly releases)
- Type: Controlled vocabulary
- Domain: Molecular biology
- Developer: GO Consortium
- Funding: NIH (grants)
- Availability
  - Publicly available: Yes
  - Repositories: UMLS / OBO / NCBO BioPortal
- URL: http://www.geneontology.org/
Gene Ontology Characteristics (2)

◆ Number of
  ● Terms: 2.15 per concept

◆ Major organizing principles:
  ● 3 major hierarchies
    ■ Molecular function
    ■ Cellular component
    ■ Biological process
  ● Relations (within hierarchies): isa, part_of, regulates
  ● No relations between concepts across hierarchies

◆ Formalism: OBO format
Gene Ontology  Top level (MF)

- all : all [250418 gene products]
  - GO:0008150 : biological_process [166605 gene products]
  - GO:0005575 : cellular_component [169814 gene products]
  - GO:0003674 : molecular_function [168558 gene products]
    - GO:0016209 : antioxidant activity [566 gene products]
    - GO:0015457 : auxiliary transport protein activity [161 gene products]
    - GO:0005488 : binding [46697 gene products]
    - GO:0003824 : catalytic activity [51856 gene products]
    - GO:0030188 : chaperone regulator activity [73 gene products]
    - GO:0042056 : chemoattractant activity [14 gene products]
    - GO:0045499 : chemorepellent activity [9 gene products]
    - GO:0030234 : enzyme regulator activity [2370 gene products]
    - GO:0016530 : metallochaperone activity [47 gene products]
    - GO:0060089 : molecular transducer activity [7873 gene products]
    - GO:0003774 : motor activity [527 gene products]
    - GO:0045735 : nutrient reservoir activity [49 gene products]
    - GO:0031386 : protein tag [18 gene products]
    - GO:0005198 : structural molecule activity [4324 gene products]
    - GO:0030528 : transcription regulator activity [10429 gene products]
    - GO:0045182 : translation regulator activity [893 gene products]
    - GO:0005215 : transporter activity [10583 gene products]
Gene Ontology  Top level (CC)

- all: all [250418 gene products]
  - GO:0008150: biological_process [166605 gene products]
  - GO:0005575: cellular_component [169814 gene products]
    - GO:0005623: cell [111086 gene products]
    - GO:0044464: cell part [111049 gene products]
    - GO:0031975: envelope [3316 gene products]
    - GO:0031012: extracellular matrix [573 gene products]
    - GO:0044420: extracellular matrix part [292 gene products]
    - GO:0005576: extracellular region [5001 gene products]
    - GO:0044421: extracellular region part [3365 gene products]
    - GO:0032991: macromolecular complex [14668 gene products]
    - GO:0031974: membrane-enclosed lumen [5290 gene products]
    - GO:0043226: organelle [79653 gene products]
    - GO:0044422: organelle part [16645 gene products]
    - GO:0055044: symplast [3 gene products]
    - GO:0045202: synapse [454 gene products]
    - GO:0044456: synapse part [210 gene products]
    - GO:0019012: virion [227 gene products]
    - GO:0044423: virion part [186 gene products]
    - GO:0003674: molecular_function [168558 gene products]
Gene Ontology  Top level (BP)

- all : all [250418 gene products]
  - GO:0008150 : biological_process [166605 gene products]
    - GO:0026610 : biological adhesion [1586 gene products]
    - GO:0065007 : biological regulation [31031 gene products]
    - GO:0001906 : cell killing [177 gene products]
    - GO:0009987 : cellular process [79087 gene products]
    - GO:0032502 : developmental process [19678 gene products]
    - GO:0051234 : establishment of localization [15270 gene products]
    - GO:0040007 : growth [4139 gene products]
    - GO:0002376 : immune system process [2517 gene products]
    - GO:0051179 : localization [17811 gene products]
    - GO:0040011 : locomotion [1251 gene products]
    - GO:0008152 : metabolic process [61127 gene products]
    - GO:0051704 : multi-organism process [4780 gene products]
    - GO:0032501 : multicellular organismal process [20567 gene products]
    - GO:0048519 : negative regulation of biological process [5037 gene products]
    - GO:0043473 : pigmentation [235 gene products]
    - GO:0048518 : positive regulation of biological process [6585 gene products]
    - GO:0050789 : regulation of biological process [28645 gene products]
    - GO:0000003 : reproduction [6343 gene products]
    - GO:0022414 : reproductive process [3535 gene products]
    - GO:0050896 : response to stimulus [16487 gene products]
    - GO:0048511 : rhythmic process [404 gene products]
    - GO:0016032 : viral reproduction [536 gene products]
Gene Ontology

- all : all [250418 gene products]
  - GO:0003674 : molecular_function [168558 gene products]
  - GO:0003824 : catalytic activity [51856 gene products]
  - GO:0016740 : transferase activity [15763 gene products]
    - GO:0016772 : transferase activity, transferring phosphorus-containing groups
  - GO:0016301 : kinase activity [6093 gene products]
    - GO:0004672 : protein kinase activity [3504 gene products]
      - GO:0004712 : protein serine/threonine/tyrosine kinase activity
  - GO:0004708 : MAP kinase kinase activity
    - GO:0016773 : phosphotransferase activity, alcohol group as acceptor
    - GO:0004672 : protein kinase activity [3504 gene products]
      - GO:0004712 : protein serine/threonine/tyrosine kinase activity
    - GO:0004708 : MAP kinase kinase activity

Lister Hill National Center for Biomedical Communications
RxNorm
RxNorm Characteristics (1)

- Current version: June 2, 2007 (monthly releases)
- Type: Controlled terminology
- Domain: Drug names
- Developer: NLM
- Funding: NLM
- Availability
  - Publicly available: Yes*
  - Repositories: UMLS
- URL: http://www.nlm.nih.gov/research/umls/rxnorm/
RxNorm Characteristics (2)

◆ Number of
  • Concepts: 93k
  • Terms: 105k

◆ Major organizing principles:
  • Generic vs. brand
  • Combinations of Ingredient / Form / Dose
  • No hierarchical structure
  • Links to all major US drug information sources
  • No clinical information

◆ Formalism: UMLS RRF format
RxNorm Normalized form

- **Strength**: 4mg/ml
- **Ingredient**: Fluoxetine
- **Dose form**: Oral Solution

Semantic clinical drug component

Semantic clinical drug form

Semantic clinical drug
Rx Norm  Generic vs. Brand

◆ Generic
  ● Ingredient (IN)
  ● Clinical drug form (SCDF)
  ● Clinical drug component (SCDC)
  ● Clinical drug (SCD)

◆ Brand
  ● Brand name (BN)
  ● Branded drug form (SBDF)
  ● Branded drug component (SBDC)
  ● Branded drug (SBD)

tradename_of
RxNorm Relations among drug entities
Medical Subject Headings (MeSH)
MeSH Characteristics (1)

- Current version: 2008 (yearly releases)
- Type: Thesaurus / Controlled vocabulary
- Domain: Biomedicine
- Developer: NLM
- Funding: NLM (Library Operations)

Availability
- Publicly available: Yes
- Repositories: UMLS / NCBO BioPortal

URL: http://www.nlm.nih.gov/mesh/
MeSH Characteristics (2)

◆ Number of
  ● Terms: 7.5 per descriptor

◆ Major organizing principles:
  ● Descriptor + entry terms
    (also: Qualifiers, Supplementary concepts)
  ● Thesaurus relations (RB/RN/RO)

◆ Formalism: Thesaurus / Proprietary XML DTD
MeSH Top level

1. Anatomy [A]
2. Organisms [B]
3. Diseases [C]
4. Chemicals and Drugs [D]
5. Analytical, Diagnostic and Therapeutic Techniques and Equipment [E]
6. Psychiatry and Psychology [F]
7. Biological Sciences [G]
8. Natural Sciences [H]
9. Anthropology, Education, Sociology and Social Phenomena [I]
10. Technology, Industry, Agriculture [J]
11. Humanities [K]
12. Information Science [L]
13. Named Groups [M]
14. Health Care [N]
15. Publication Characteristics [V]
16. Geographicals [Z]
# MeSH Example (terms)

<table>
<thead>
<tr>
<th>MeSH Heading</th>
<th>Hydrocortisone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree Number</td>
<td>D04.808.745.745.654.600</td>
</tr>
<tr>
<td>Tree Number</td>
<td>D06.472.040.585.353.476</td>
</tr>
<tr>
<td>Tree Number</td>
<td>D06.472.040.585.478.392</td>
</tr>
<tr>
<td>Scope Note</td>
<td>The main glucocorticoid secreted by the ADRENAL CORTEX. Its synthetic counterpart is used, either as an injection or topically, in the treatment of inflammation, allergy, collagen diseases, asthma, adrenocortical deficiency, shock, and some neoplastic conditions.</td>
</tr>
<tr>
<td>Entry Term</td>
<td>11-Epicortisol</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Cortifair</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Cortisol</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Cortril</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Epicortisol</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Hydrocortisone, (11 alpha)-Isomer</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Hydrocortisone, (9 beta,10 alpha,11 alpha)-Isomer</td>
</tr>
</tbody>
</table>
MeSH Example (hierarchies)

Chemicals and Drugs

Polycyclic Compounds

Steroids

Pregnanes

Pregnenes

Pregnenediones

Hormones, Hormone Substitutes, and Hormone Antagonists

Hormones

Adrenal Cortex Hormones

Hydroxycorticosteroids

11-Hydroxycorticosteroids

11-Hydroxycorticosteroids

Hydrocortisone
NCI Thesaurus
NCI thesaurus Characteristics (1)

- Current version: 08.04d (~monthly releases)
- Type: Controlled terminology / ontology
- Domain: Cancer
- Developer: NCI Center for Bioinformatics
- Funding: NCI
- Availability
  - Publicly available: Yes
  - Repositories: UMLS / OBO / NCBO BioPortal
- URL: http://nciterms.nci.nih.gov/
NCI thesaurus  Characteristics (2)

◆ Number of
  ● Concepts: 58,868 (2007_05E)
  ● Terms: 2.68 per concept

◆ Major organizing principles:
  ● Subsumption hierarchy
  ● Rich set of associative relationships
  ● Small proportion of defined concepts (many primitives)
  ● Links to many external resources

◆ Formalism: OWL Lite
NCI thesaurus  Top level

NCI_Thesaurus Taxonomy

- Abnormal Cell
- Activity
- Anatomic Structure, System, or Substance
- Biochemical Pathway
- Biological Process
- Chemotherapy Regimen or Agent Combination
- Conceptual Entity
- Diagnostic, Therapeutic, and Research Equipment
- Diagnostic or Prognostic Factor
- Disease, Disorder or Finding
- Drug, Food, Chemical or Biomedical Material
- Experimental Organism Anatomical Concept
- Experimental Organism Diagnosis
- Gene
- Gene Product
- Molecular Abnormality
- NCI Administrative Concept
- Organism
- Property or Attribute
- Retired Concept
NCI thesaurus  Example

Concept Details

URI: http://htterms.nlm.nih.gov:80/NCIBrowser/Con...dictionary=NCI_Thesaurus&code=C2919
Version: June 2007 (07.06d)

Prostate Adenocarcinoma

Identifiers:
- name: Prostate_Adenocarcinoma
- code: C2919

Relationships to other concepts:
- Disease_Has_Finding: Invasive Lesion
- Disease_Has_Abnormal_Cell: Adenocarcinoma Cell
- Disease_Has_Normal_Tissue_Origin: Prostatic Epithelium
- Disease_MAY_Have_Finding: Serum Prostate Specific Antigen Increased
- Disease_Has_Finding: Carcinomatous Component Present
- Disease_Excludes_Abnormal_Cell: Neoplastic Smooth Muscle Cell
- Disease_Excludes_Abnormal_Cell: Malignant Squamous Cell
- Disease_Has_Parent_Anatomic_Site: Prostate Gland
- Disease_Has_Associated_Anatomic_Site: Male Reproductive System
- Disease_Excludes_Abnormal_Cell: Malignant Stromal Cell
- Disease_Has_Associated_Anatomic_Site: Prostate Gland
- Disease_Has_Normal_Cell_Origin: Epithelial Cell

Superconcepts:
- Adenocarcinoma
- Common Carcinoma
- Invasive Prostate Carcinoma

Subconcepts:
- Acinar Prostate Adenocarcinoma
- Metastatic Prostatic Adenocarcinoma
- Moderately Differentiated Prostate Adenocarcinoma
- Poorly Differentiated Prostate Adenocarcinoma
- Prostate Adenocarcinoma with Focal Neuroendocrine Differentiation
- Prostate Ductal Adenocarcinoma
- Stage III Prostate Adenocarcinoma
- Stage II Prostate Adenocarcinoma
- Stage I Prostate Adenocarcinoma
- Well Differentiated Prostate Adenocarcinoma

Information about this concept:

- Definition
- Synonym with source data
- Synonym with source data
- Synonym with source data
- Preferred_Name
- Semantic_Type
- Synonym
- Synonym
- Unified Medical Language System Concept Identifier
Unified Medical Language System (UMLS)
UMLS Characteristics (1)

- Current version: 2008AA (2-3 annual releases)
- Type: Terminology integration system
- Domain: Biomedicine
- Developer: NLM
- Funding: NLM (intramural)
- Availability
  - Publicly available: Yes* (cost-free license required)
  - Repositories: UMLS
- URL: http://umlsks.nlm.nih.gov/
UMLS Characteristics (2)

- Number of
  - Concepts: 1.5M (2008AA)
  - Terms: ~6M

- Major organizing principles (Metathesaurus):
  - Concept orientation
  - Source transparency
  - Multi-lingual through translation

- Formalism: Proprietary format (RRF)
UMLS Integrating subdomains

Clinical repositories

Genetic knowledge bases

Other subdomains

OMIM

SNOMED CT

Biomedical literature

UMLS

MeSH

NCBI Taxonomy

Model organisms

GO

FMA

Genome annotations

Anatomy

Lister Hill National Center for Biomedical Communications
Addison’s Disease: Concept

Addison’s Disease

Disease or Syndrome

ADRENAL INSUFFICIENCY (ADDISON'S DISEASE)
ADRENOCORTICAL INSUFFICIENCY, PRIMARY FAILURE
Hypoadrenalisms, Primary
Melasma addisonii
Primary adrenal deficiency
Asthenia pigmentosa
Bronzed disease
Insufficiency, adrenal primary
Primary adrenocortical insufficiency
Addison's, disease

Maladie d'Addison - French
Addison-Krankheit - German
Morbo di Addison - Italian
Doença de Addison - Portuguese
АДДИСОНОВА БОЛЕЗНЬ - Russian
アジソン病 - Japanese

An adrenal disease characterized by the progressive destruction of the adrenal cortex, resulting in insufficient production of aldosterone and hydrocortisone. Clinical symptoms include anorexia; nausea; weight loss; muscle weakness; and hyperpigmentation of the skin due to increase in circulating levels of ACTH precursor hormone which stimulates melanocytes.

SNOMED CT
SNOMED Intl
MeSH
MedDRA

C0001403
Metathesaurus Concepts (2007AB)

- **Concept** (~1.4M) CUI
  - Set of synonymous concept names
- **Term** (~5.3M) LUI
  - Set of normalized names
- **String** (~5.9M) SUI
  - Distinct concept name
- **Atom** (~7.2M) AUI
  - Concept name in a given source

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A0066000</td>
<td>Headache (MeSH)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0065992</td>
<td>Headache (ICD-10)</td>
<td>S0046854</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0066007</td>
<td>Headaches (MedDRA)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A12003304</td>
<td>Headaches (OMIM)</td>
<td>S0046855</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A0540936</td>
<td>Cephalodynia (MeSH)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Lister Hill National Center for Biomedical Communications
### Recap

<table>
<thead>
<tr>
<th>Name</th>
<th>Scope</th>
<th># concepts</th>
<th>Median</th>
<th>Subs. Hier</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNOMED CT</td>
<td>Clinical medicine (patient records)</td>
<td>310,314</td>
<td>2</td>
<td>yes</td>
<td>July 31, 2007</td>
</tr>
<tr>
<td>LOINC</td>
<td>Clinical observations and laboratory tests</td>
<td>46,406</td>
<td>3</td>
<td>no</td>
<td>Version 2.21</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(no “natural language” names)</td>
</tr>
<tr>
<td>FMA</td>
<td>Human anatomical structures</td>
<td>~72,000</td>
<td>?</td>
<td>yes</td>
<td>(not yet in the UMLS)</td>
</tr>
<tr>
<td>RxNorm</td>
<td>Standard names for prescription drugs</td>
<td>93,426</td>
<td>1</td>
<td>no</td>
<td>Aug. 31, 2007</td>
</tr>
<tr>
<td>NCI Thesaurus</td>
<td>Cancer research, clinical care, public information</td>
<td>58,868</td>
<td>2</td>
<td>yes</td>
<td>2007_05E</td>
</tr>
<tr>
<td>ICD-10</td>
<td>Diseases and conditions (health statistics)</td>
<td>12,318</td>
<td>1</td>
<td>no</td>
<td>1998 (tabular)</td>
</tr>
<tr>
<td>MeSH</td>
<td>Biomedicine (descriptors for indexing the literature)</td>
<td>24,767</td>
<td>5</td>
<td>no</td>
<td>Aug. 27, 2007</td>
</tr>
<tr>
<td>UMLS .</td>
<td>Terminology integration in the life sciences</td>
<td>1,4 M</td>
<td>2</td>
<td>n/a</td>
<td>2007AC (English only)</td>
</tr>
<tr>
<td>Day</td>
<td>Topic</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>----------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monday, June 9</td>
<td>Introduction to Biomedical Ontologies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Design Principles, Formalisms and Tools for Biomedical Ontologies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biomedical Ontologies - Content and structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Biomedical Ontologies - Function</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Contrasting Biomedical Ontologies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Extending Biomedical Ontologies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Using Biomedical Ontologies for Data Integration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Biomedical Ontologies in Action

A Functional Perspective on Biomedical Ontologies

Olivier Bodenreider
Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA
Overview

◆ Functional perspective
  ● What are they for (vs. what are they)?
◆ “High-impact” biomedical ontologies
◆ 3 major categories of use
  ● Knowledge management (indexing and retrieval of data and information, access to information, mapping among ontologies)
  ● Data integration, exchange and semantic interoperability
  ● Decision support and reasoning (data selection and aggregation, decision support, natural language processing applications, knowledge discovery).
Knowledge management
Knowledge management

Annotating data and resources
Terminology in ontology

- Ontology as a source of vocabulary
  - List of names for the entities in the ontology (ontology vs. terminology)

- Most ontologies have some sort of terminological component
  - Exceptions: GALEN, LOINC

- Not all surface forms represented
  - Often insufficient for NLP applications
  - Large variation in number of terms per concept across ontologies
Annotating data

◆ Gene Ontology
  • Functional annotation of gene products in several dozen model organisms

◆ Various communities use the same controlled vocabularies

◆ Enabling comparisons across model organisms

◆ Annotations
  • Assigned manually by curators
  • Inferred automatically (e.g., from sequence similarity)
### GO Annotations for Aldh2 (mouse)

<table>
<thead>
<tr>
<th>Category</th>
<th>Classification Term</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecular Function</td>
<td>aldehyde_dehydrogenase (NAD) activity</td>
<td>IEA</td>
</tr>
<tr>
<td>Molecular Function</td>
<td>oxidoreductase activity</td>
<td>IEA</td>
</tr>
<tr>
<td>Molecular Function</td>
<td>oxidoreductase activity</td>
<td>IEA</td>
</tr>
<tr>
<td>Cellular Component</td>
<td>mitochondrion</td>
<td>IDA</td>
</tr>
<tr>
<td>Biological Process</td>
<td>metabolic process</td>
<td>IEA</td>
</tr>
<tr>
<td>Biological Process</td>
<td>oxidation reduction</td>
<td>IEA</td>
</tr>
</tbody>
</table>

http://www.informatics.jax.org/
GO ALD4 in Yeast

All ALD4 GO evidence and references

View Computational GO annotations for ALD4

- aldehyde dehydrogenase (NAD) activity (IDA, IMP, ISS)
- aldehyde dehydrogenase [NAD(P)+] activity (IDA)
- ethanol metabolic process (IMP)
- mitochondrial nucleoid (IDA)
- mitochondrion (IMP, ISS)
- mitochondrion (IDA)

http://db.yeastgenome.org/
## GO Annotations for ALDH2 (Human)

<table>
<thead>
<tr>
<th>GO:0016491</th>
<th>oxidoreductase activity</th>
<th>interpro</th>
<th>EA</th>
<th>IPR015590</th>
<th>UniProt 9606</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO:0016491</td>
<td>oxidoreductase activity</td>
<td>interpro</td>
<td>EA</td>
<td>IPR016160</td>
<td>UniProt 9606</td>
</tr>
<tr>
<td>GO:0016491</td>
<td>oxidoreductase activity</td>
<td>interpro</td>
<td>EA</td>
<td>IPR016162</td>
<td>UniProt 9606</td>
</tr>
<tr>
<td>GO:0016491</td>
<td>oxidoreductase activity</td>
<td>interpro</td>
<td>EA</td>
<td>IPR016161</td>
<td>UniProt 9606</td>
</tr>
<tr>
<td>GO:0016491</td>
<td>oxidoreductase activity</td>
<td>spkw</td>
<td>EA</td>
<td>KW-0560</td>
<td>UniProt 9606</td>
</tr>
<tr>
<td>GO:0004029</td>
<td>aldehyde dehydrogenase (NAD) activity</td>
<td>1306115</td>
<td>TAS</td>
<td>PINC 9606</td>
<td></td>
</tr>
<tr>
<td>GO:0004030</td>
<td>aldehyde dehydrogenase [NAD(P)+] activity</td>
<td>8903321</td>
<td>TAS</td>
<td>PINC 9606</td>
<td></td>
</tr>
<tr>
<td>GO:0009055</td>
<td>electron carrier activity</td>
<td>8903321</td>
<td>TAS</td>
<td>UniProt 9606</td>
<td></td>
</tr>
<tr>
<td>GO:0004029</td>
<td>aldehyde dehydrogenase (NAD) activity</td>
<td>enzyme</td>
<td>EA</td>
<td>1.2.1.3</td>
<td>UniProt 9606</td>
</tr>
</tbody>
</table>

[http://www.ebi.ac.uk/GOA/](http://www.ebi.ac.uk/GOA/)
Indexing the biomedical literature

◆ MeSH
  • Used for indexing and retrieval of the biomedical literature (MEDLINE)

◆ Indexing
  • Performed manually by human indexers
    - With help of semi-automatic systems (suggestions)
      e.g., Indexing Initiative at NLM
  • Automatic indexing systems
Free cortisol in sepsis and septic shock.


Department of Intensive Care, Kuopio University Hospital, PL 16222 Kuopio, Finland. Stepani.Bendel@kuh.fi

BACKGROUND: Severe sepsis activates the hypothalamic-pituitary axis, increasing cortisol production. In some studies, hydrocortisone substitution based on an adrenocorticotropic hormone-stimulation test or baseline cortisol measurement has improved outcome. Because only the free fraction of cortisol is active, measurement of free cortisol may be more important than total cortisol in critically ill patients. We measured total and free cortisol in patients with severe sepsis and related the concentrations to outcome. METHODS: In a prospective study, severe sepsis was defined according the American College of Chest Physicians/Society of Critical Care Medicine criteria. Blood samples were drawn within 24 h of study entry. Serum cortisol was analyzed by electrochemiluminescence immunoassay. The Coolens method was used for calculating serum free cortisol concentrations. RESULTS: Blood samples were collected from 125 patients, of whom 62 had severe sepsis and 63 septic shock. Hospital mortality was 21%. Calculated free serum cortisol correlated well with serum total cortisol ($r = 0.90, P < 0.001$). There was no difference in the total cortisol concentrations in patients with sepsis and septic shock ($728 +/- 386$ nmol/L vs $793 +/- 439$ nmol/L, $P = 0.44$). Nonsurvivors had higher calculated serum free ($209 +/- 151$ nmol/L) and total ($980 +/- 458$ nmol/L) cortisol concentrations than survivors ($119 +/- 111$ nmol/L, $P = 0.002$, and $704 +/- 383$ nmol/L, $P = 0.002$). Depending on the definition, the incidence of adrenal insufficiency varied from 8% to 54%.

CONCLUSIONS: Clinically, calculation of free cortisol does not provide essential information for identification of patients who would benefit from corticoid treatment in severe sepsis and septic shock.
MeSH MEDLINE indexing

MeSH Terms:
- Adrenal Cortex Function Tests
- Adrenal Insufficiency/blood*
- Adrenal Insufficiency/drug therapy
- Adrenal Insufficiency/mortality
- Adult
- Biological Markers/blood
- Female
- Finland/epidemiology
- Hospital Mortality
- Humans
- Hydrocortisone/blood*
- Hydrocortisone/therapeutic use
- Kaplan-Meiers Estimate

- Male
- Predictive Value of Tests
- Prospective Studies
- Sepsis/blood*
- Sepsis/drug therapy
- Sepsis/mortality
- Severity of Illness Index
- Shock, Septic/blood*
- Shock, Septic/drug therapy
- Shock, Septic/mortality
- Treatment Outcome

Substances:
- Biological Markers
- Hydrocortisone
Replacement therapy for Addison's disease: recent developments.

Lovás K, Husebye ES.

University of Bergen, Institute of Medicine, Section of Endocrinology, 5021 Bergen, Norway.
Kristian.lovas@helse-bergen.no

BACKGROUND: The hormone deficiencies in Addison's disease (primary adrenal insufficiency) are conventionally treated with oral glucocorticoid and mineralocorticoid replacement but the available therapies do not restore the physiological hormone levels and biorhythm. Despite such treatment these patients self-report impaired health-related quality of life (HRQoL) and recent research has indicated increased mortality. OBJECTIVE/METHODS: We review the literature and recent developments in replacement therapy. RESULTS/CONCLUSION: Patients with Addison's disease require mineralocorticoid replacement, i.e., fludrocortisone 0.05 - 0.20 mg once daily. Starting doses of glucocorticoids should be 15 - 20 mg for hydrocortisone or 20 - 30 mg for cortisol acetate, divided into two or three doses, and preferentially weight-adjusted. There are indications that the synthetic glucocorticoids have undesirable metabolic long-term effects, which make them less suitable as first-line treatment. Timed-release hydrocortisone tablets and continuous subcutaneous hydrocortisone infusion are promising new treatment modalities. Studies of replacement with the adrenal androgen dehydroepiandrosterone (DHEA) in adrenal failure have shown inconsistent benefit on HRQoL. DHEA, or possibly testosterone replacement is likely to be beneficial for selected groups of patients with Addison's disease but this remains to be shown. We here give our opinion of the best treatment and future direction of research in this area.
MeSH MEDLINE indexing

MeSH Terms:
- Addison Disease/blood
- Addison Disease/drug therapy*
- Androgens/administration & dosage*
- Androgens/therapeutic use
- Dosage Forms
- Drug Administration Routes
- Drug Administration Schedule
- Glucocorticoids/administration & dosage*
- Glucocorticoids/adverse effects
- Glucocorticoids/blood
- Glucocorticoids/deficiency
- Hormone Replacement Therapy*
- Humans
- Mineralocorticoids/administration & dosage*
- Mineralocorticoids/adverse effects
- Mineralocorticoids/blood
- Mineralocorticoids/deficiency
- Quality of Life
- Treatment Outcome

Substances:
- Androgens
- Dosage Forms
- Glucocorticoids
- Mineralocorticoids
ICD9-CM  Coding clinical data

◆ ICD9-CM
  • Used for coding clinical data e.g., for billing purposes

◆ Other uses of ICD
  • Morbidity and mortality reporting worldwide
Knowledge management

Accessing biomedical information
Resources for biomedical search engines

◆ Synonyms
◆ Hierarchical relations
◆ High-level categorization
◆ Co-occurrence information
◆ Translation
MeSH “synonyms” MEDLINE retrieval

- MeSH entry terms
  - Used as equivalent terms for retrieval purposes
  - Not always synonymous
- Increase recall without hurting precision

<table>
<thead>
<tr>
<th>MeSH Heading</th>
<th>Addison Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Entry Term</td>
<td>Addison's Disease</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Primary Adrenal Insufficiency</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Primary Adrenocortical Insufficiency</td>
</tr>
</tbody>
</table>
MeSH “synonyms” MEDLINE retrieval

[Image of PubMed search interface with the query: Primary Hypoadrenalism]
MeSH hierarchies  MEDLINE retrieval

◆ MeSH “explosion”
  • Search for a given MeSH term and all its descendants
  • A search on Adrenal insufficiency also retrieves articles indexed with Addison disease

   Free cortisol in sepsis and septic shock.
   PMID: 18499615 [PubMed - indexed for MEDLINE]

2: Luboshitsky R, Quptt O.

   Corticosteroids for septic shock.
   PMID: 18467975 [PubMed - indexed for MEDLINE]

12: Lovás K, Husebye ES.

   Replacement therapy for Addison's disease: recent developments.
   PMID: 18363515 [PubMed - indexed for MEDLINE]
Co-indexing

http://www.gopubmed.com/

cox-2

Lister Hill National Center for Biomedical Communications

Top categories
- Chemicals and Drugs [992]
- Cyclooxygenase 2 [517]
- Cyclooxygenase 2 Inhibitors [289]
- Prostaglandins [398]
- Prostaglandin-Endoperoxide Synthases [683]
- NF-kappa B [138]
- RNA, Messenger [222]
- Anti-Inflammatory Agents [414]
- more
- Biological Process [851]
- Cyclooxygenase Pathway [305]
- more
- Biological Sciences [960]
- Up-Regulation [166]
- more
- Diseases [781]
- Inflammation [192]
- more
- Organisms [398]
- Techniques and Equipment [809]
- molecular function [483]
- Anatomy [778]
- Named Groups [285]
- cellular component [307]
- Natural Sciences [661]
- Technology, Industry, Agriculture [147]
- Psychiatry and Psychology [386]
Knowledge management

Mapping across biomedical ontologies
Reusing information

- Clinical information coded with SNOMED CT
  - Mapped to ICD9-CM and CPT for billing purposes
  - Mapped to ICD-O for epidemiology purposes
- Existing mapping tables created by terminology developers as an incentive to use SNOMED CT
Reusing tools

- For noun phrases extracted from medical texts, map to UMLS concepts
- Then, select from the MeSH vocabulary the concepts that are the most closely related to the original concepts

[Aronson & al., AMIA, 2000]
Terminology integration systems

- Terminology integration systems (UMLS, RxNorm) help bridge across vocabularies

- Uses
  - Information integration
  - Ontology alignment
  - Medication reconciliation
Integrating subdomains

Clinical repositories

Genetic knowledge bases

Other subdomains

SNOMED CT

OMIM

Biomedical literature

UMLS

MeSH

NCBI Taxonomy

Model organisms

GO

FMA

Genome annotations

Anatomy

Lister Hill National Center for Biomedical Communications
Integrating subdomains

- Clinical repositories
- Genetic knowledge bases
- Other subdomains
- Biomedical literature
- Model organisms
- Genome annotations
- Anatomy
Trans-namespaece integration

- Addison's disease (363732003)
- Clinical repositories
- Genetic knowledge bases
- Other subdomains
- SNOMED CT
- OMIM
- Other subdomains
- UMLS
- C0001403
- GO
- FMA
- MESH
- Addision Disease (D000224)
- MeSH
- Biomedical literature
- Genome annotations
- Model organisms
- Anatomy
- NCBI Taxonomy
- Lister Hill National Center for Biomedical Communications
Data integration, exchange and semantic interoperability
Data integration, exchange and semantic interoperability

Information exchange and semantic operability
“Standards”

◆ Ontologies help standardize patients data
  • Facilitate the exchange of data across institutions
  • Help connect “islands of data” (silos)

◆ LOINC
  • Exchange of laboratory data
  • In conjunction with HL7 messaging
Semantic interoperability projects  BRIDG

- Biomedical Research Integrated Domain Group
  - Information model for clinical research
  - Interoperability between clinical trials information systems
  - Ontologies provide value sets to the information model
Semantic interoperability projects  CDA

- Clinical Document Architecture (CDA R2)
  - Formal representation of clinical statements
    - Clinical observations
    - Medication administration
    - Adverse events
  - Associate an information model (HL7 RIM) with terminologies (LOINC, SNOMED CT, RxNorm)
Semantic interoperability projects caCORE

- Cancer Common Ontologic Representation Environment
  - Infrastructure developed to support an interoperable biomedical information system for cancer research
  - Uses the NCI Thesaurus as a component
Data integration, exchange and semantic interoperability

Information and data integration
Approaches to data integration

◆ Warehousing
  ● Sources to be integrated are transformed into a common format and converted to a common vocabulary
  ● Normalization through ontologies (e.g., GO annotations)

◆ Mediation
  ● Local schema (of the sources)
  ● Global schema (in reference to which the queries are made)
  ● Ontologies help define the global schema and map between local and global schemas (OntoFusion, ARIANE)
Ontologies and integration

- Terminology integration systems help bridge across terminologies and the domains they represent
- Mappings across ontologies enable the integration of namespaces in the Semantic Web
Trans-namesapce integration

Addison's disease
(363732003)

Clinical repositories

Genetic knowledge bases

Other subdomains

SNOMED CT

OMIM

UMLS

C0001403

Biomedical literature

Addison Disease (D000224)

Genome annotations

MeSH

NCBI Taxonomy

FMA

GO

Model organisms

Anatomy

Lister Hill National Center for Biomedical Communications
Decision support and reasoning
Data selection

◆ The structure of biomedical ontologies helps define groups of values from a high-level value
  ● Vs. enumerating all possible values
◆ Useful for data selection in clinical studies
◆ ICD is used pervasively for this purpose
  ● E.g., Study on supraventricular tachycardia (SVT), based on 2 high-level ICD codes

◆ Similarity with the definition of value sets for use in the information model
Data aggregation

◆ Ontologies help partition/aggregate data in data analysis
  ● Clinical studies: Study a variable in groups of patients corresponding to the top level categories in ICD
  ● Biology studies: Functional characterization of gene expression signatures with high-level concepts from the Gene Ontology
    ■ Recent trend: co-clustering
Decision support

◆ Clinical decision support
  - Ontologies help normalize the vocabulary and increase the recall of rules
  - Ontologies provide some domain knowledge and make it possible to create high-level rules (e.g., for a class of drugs rather than for each drug in the class)

◆ Other forms of decision support
  - Based on automatic reasoning services for OWL ontologies (e.g., grading gliomas with NCIt)
Natural language processing applications

- Ontologies provide background domain knowledge for NLP applications
  - Question answering
  - Document summarization
  - Literature-based discovery
- The UMLS is often used, but other specific resources have been developed
Knowledge discovery

- By standardizing the vocabulary in a given domain, ontologies are enabling resources for knowledge discovery through data mining.
- Less frequently, the structure of the ontology is leveraged by data mining algorithms.
- Example of available datasets:
  - ICD-coded clinical data (in conjunction with non-clinical information, e.g., environmental data)
  - Annotation of gene products to the GO (function prediction)
Barriers to usability of biomedical ontologies
Availability

- Many ontologies are freely available
- The UMLS is freely available for research purposes
  - Cost-free license required
- Licensing issues can be tricky
  - SNOMED CT is freely available in member countries of the IHTSDO
- Being freely available
  - Is a requirement for the Open Biomedical Ontologies (OBO)
  - Is a de facto prerequisite for Semantic Web applications
Discoverability

◆ Ontology repositories
  - UMLS: 143 source vocabularies
    (biased towards healthcare applications)
  - NCBO BioPortal: ~100 ontologies
    (biased towards biological applications)
  - Limited overlap between the two repositories

◆ Need for discovery services
Formalism

♦ Several major formalism
  • Web Ontology Language (OWL) – NCI Thesaurus
  • OBO format – most OBO ontologies
  • UMLS Rich Release Format (RRF) – UMLS, RxNorm

♦ Conversion mechanisms
  • OBO to OWL
  • LexGrid (import/export to LexGrid internal format)
Ontology integration

◆ *Post hoc* integration, form the bottom up
  - UMLS approach
  - Integrates ontologies “as is”, including legacy ontologies
  - Facilitates the integration of the corresponding datasets

◆ Coordinated development of ontologies
  - OBO Foundry approach
  - Ensures consistency *ab initio*
  - Excludes legacy ontologies
Quality

◆ Quality assurance in ontologies is still imperfectly defined
  ● Difficult to define outside a use case or application
◆ Several approaches to evaluating quality
  ● Collaboratively, by users (Web 2.0 approach)
    ■ Marginal notes enabled by BioPortal
  ● Centrally, by experts
    ■ OBO Foundry approach
◆ Important factors besides quality
  ● Governance
  ● Installed base / Community of practice
<table>
<thead>
<tr>
<th>Day</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday, June 9</td>
<td>Introduction to Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Design Principles, Formalisms and Tools for Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Biomedical Ontologies - Content and structure - Function</td>
</tr>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Searching and Analyzing Biomedical Concepts</td>
</tr>
<tr>
<td></td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Extending Biomedical Ontologies</td>
</tr>
<tr>
<td></td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>
Interfaces to Biomedical Ontologies

Olivier Bodenreider
Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Overview

◆ Graphical interfaces
  • UMLS Knowledge Source Server
  • NCBO BioPortal
  • NCI Thesaurus
  • MeSH browser
  • Foundational Model of Anatomy (FMA) Explorer
  • Gene Ontology AmiGO
  • ICD-10 online
  • RxNav (RxNorm)
  • […]

◆ Application Programming Interfaces
Graphical interfaces
UMLS Knowledge Source Server

http://umlsks.nlm.nih.gov/

Lister Hill National Center for Biomedical Communications
UMLS Semantic Navigator

Lister Hill National Center for Biomedical Communications

NCBO BioPortal

**Ontologies**

<table>
<thead>
<tr>
<th>Name</th>
<th>Format</th>
<th>Current Version</th>
<th>Content Location</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>African Traditional Medicine</td>
<td>OBO</td>
<td>1.0.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Amino Acid</td>
<td>OWL Full</td>
<td>1.2</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Amphibian gross anatomy</td>
<td>OBO</td>
<td>1.7</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Animal natural history and life history</td>
<td>Protege</td>
<td>See Remote Site</td>
<td>Remote</td>
<td></td>
</tr>
<tr>
<td>Basic Vertebrate Anatomy</td>
<td>OWL Full</td>
<td>1.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Biological imaging methods</td>
<td>OBO</td>
<td>1.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Biological process</td>
<td>OBO</td>
<td>1.208</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>Biomedical Resource Ontology</td>
<td>OWL Lite</td>
<td>1.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
<tr>
<td>BIRN Lex</td>
<td>OWL DL</td>
<td>1.3.1</td>
<td>NCBO Library</td>
<td></td>
</tr>
</tbody>
</table>


Lister Hill National Center for Biomedical Communications 215
NCI Thesaurus (EVS Server)

Prostate Adenocarcinoma

Identifiers:
- name: Prostate_Adenocarcinoma
- code: C2019

Relationships to other concepts:
- Disease_Has_Normal_Tissue_Origin: Prostatic Epithelium
- Disease_Has_Abnormal_Cell: Adenocarcinoma_Cell
- Disease_Has_Associated_Anatomic_Site: Prostate Gland
- Disease_Excludes_Abnormal_Cell: Neoplastic Smooth Muscle Cell
- Disease_Has_Finding: Invasive Lesion
- Disease_Has_Primary_Anatomic_Site: Prostate Gland

http://nciterms.nci.nih.gov/NCIBrowser/SearchConcept.do

Lister Hill National Center for Biomedical Communications
MeSH Browser

<table>
<thead>
<tr>
<th>MeSH Heading</th>
<th>Prostatic Neoplasms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tree Number</td>
<td>C04.588.945.440.770</td>
</tr>
<tr>
<td>Tree Number</td>
<td>C12.294.260.750</td>
</tr>
<tr>
<td>Tree Number</td>
<td>C12.294.565.625</td>
</tr>
<tr>
<td>Tree Number</td>
<td>C12.758.409.750</td>
</tr>
<tr>
<td>Annotation</td>
<td>coordinate IM with histological type of neoplasm (IM); note PROSTATIC ADENOMA see PROSTATIC HYPERPLASIA is also available</td>
</tr>
<tr>
<td>Scope Note</td>
<td>Tumors or cancer of the PROSTATE.</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Cancer of Prostate</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Cancer of the Prostate</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Neoplasms, Prostate</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Neoplasms, Prostatic</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Prostate Cancer</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Prostate Neoplasms</td>
</tr>
<tr>
<td>Entry Term</td>
<td>Prostatic Cancer</td>
</tr>
<tr>
<td>See Also</td>
<td>Prostate-Specific Antigen</td>
</tr>
<tr>
<td>See Also</td>
<td>Prostatic Hyperplasia</td>
</tr>
<tr>
<td>Allowable Qualifiers</td>
<td>BL BS CF CH CI CL CN CO DH DI DT EC EH EM EN EP ET GE HI IM ME MI MO NU PA PC PP PS PX RA RH RI RT SC SE SU TH UL UR US VE VI</td>
</tr>
<tr>
<td>Entry Version</td>
<td>PROSTATIC NEOPL</td>
</tr>
<tr>
<td>Date of Entry</td>
<td>19990101</td>
</tr>
<tr>
<td>Unique ID</td>
<td>D011471</td>
</tr>
</tbody>
</table>


Lister Hill National Center for Biomedical Communications
Foundational Model of Anatomy

http://sig.biostr.washington.edu/projects/fm/FME/

Lister Hill National Center for Biomedical Communications
**Viral meningitis (disorder)**

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>Fully Specified Name</th>
<th>Concept Status</th>
<th>CTV3ID</th>
<th>SNOMED ID</th>
<th>Is Primitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>58170007</td>
<td>Viral meningitis (disorder)</td>
<td>Current (0)</td>
<td>xa995</td>
<td>DE-30020</td>
<td>Fully defined (0)</td>
</tr>
</tbody>
</table>

**Descriptions and Synonyms**

<table>
<thead>
<tr>
<th>Description ID</th>
<th>Term</th>
<th>Description Status</th>
<th>Description Type</th>
<th>Language Code</th>
<th>Initial Capital Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Viral meningitis</td>
<td>Current (0)</td>
<td>Preferred (1)</td>
<td>en</td>
<td>Capitalization meaningless (0)</td>
</tr>
<tr>
<td>2</td>
<td>Aseptic meningitis, viral</td>
<td>Current (0)</td>
<td>Synonym (2)</td>
<td>en</td>
<td>Capitalization meaningless (0)</td>
</tr>
<tr>
<td>3</td>
<td>Abacterial meningitis</td>
<td>Current (0)</td>
<td>Synonym (2)</td>
<td>en</td>
<td>Capitalization meaningless (0)</td>
</tr>
<tr>
<td>4</td>
<td>Aseptic meningitis</td>
<td>Non-Current (1)</td>
<td>Synonym (2)</td>
<td>en</td>
<td>Capitalization meaningless (0)</td>
</tr>
<tr>
<td>5</td>
<td>Viral meningitis NOS</td>
<td>Non-Current (1)</td>
<td>Synonym (2)</td>
<td>en</td>
<td>Capitalization meaningless (0)</td>
</tr>
<tr>
<td>6</td>
<td>Viral meningitis, NEC</td>
<td>Non-Current (1)</td>
<td>Synonym (2)</td>
<td>en</td>
<td>Capitalization meaningless (0)</td>
</tr>
</tbody>
</table>

**Parents**

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Parent Concept (This Concept IS A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Viral infections of the central nervous system (disorder)</td>
</tr>
<tr>
<td>2</td>
<td>Infective meningitis (disorder)</td>
</tr>
</tbody>
</table>

http://www.it.usyd.edu.au/~hitru/sct/A1.cgi
ICD-10

Chapter IV
Endocrine, nutritional and metabolic diseases (E00-E90)

Disorders of other endocrine glands (E20-E35)

Excludes: galactorrhoea (N54.3)
gynaecomastia (N62)

E20  Hypoparathyroidism
  Excludes: Di George’s syndrome (D55.1)
  postprocedural hypoparathyroidism (E09.2)
  tetany NOS (R29.0)
  transitory neonatal hypoparathyroidism (P71.4)

E20.0  Idiopathic hypoparathyroidism
E20.1  Pseudohypoparathyroidism
E20.8  Other hypoparathyroidism
E20.9  Hypoparathyroidism, unspecified
  Parathyroid tetany

http://www.who.int/classifications/apps/icd/icd10online/
RxNav (RxNorm)

Application Programming Interfaces
Application Programming Interface

- Expose resources in such a way that they can be integrated in programs
  - Programming “against” a resource
- Standard protocols for communication
  - Web services (SOAP, REST)
- Standard libraries for programming
- Focus on content, not message
UMLSKS Web Service API

  - Developer's Guide > Webservice Operations
- WSDL available
- API give access to all 3 knowledge sources
- Licensing issues
  - Granting ticket and Single-use tickets
ConceptIdGroup findCUIByNormString
(ConceptIdNormStringRequest request);

**Argument: ConceptIdNormStringRequest**

This class contains the arguments that further restrict the behavior of the call.

- setCasTicket (String s)
  - Single-use ticket returned by the AuthorizationPort webservice
- setRelease (String s)
  - UMLS release of interest
- setSearchString (String s)
  - input search string
- setSABs (String[] array)
  - set of source abbreviations to search
- setLanguage (String s)
  - language restriction
- setIncludeSuppressibles (boolean b)
  - true if suppressible strings are included in the search
- setCVF (long l)
  - Bit flag for the content view to search

**Return: ConceptIdGroup**

Lister Hill National Center for Biomedical Communications
Other APIs to terminology systems

- NCBO BioPortal
  http://www.bioontology.org/docs/bioportal/development/web_services.html
- OLS - Ontology Lookup Service
  http://www.ebi.ac.uk/ontology-lookup/WSDLDocumentation.do
- RxNorm
Applications based on WS APIs

◆ UMLSKS API
  - UMLSKS
    http://umlsks.nlm.nih.gov/

◆ RxNorm API
  - RxNav
  - MyMedicationList
    http://mml.nlm.nih.gov/MyMedicationList.jnlp
## Agenda

<table>
<thead>
<tr>
<th>Monday, June 9</th>
<th>Introduction to Biomedical Ontologies</th>
<th>Design Principles, Formalisms and Tools for Biomedical Ontologies</th>
<th>Biomedical Ontologies - Content and structure - Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td>Extending Biomedical Ontologies</td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>
Searching and Analyzing
Biomedical Concepts

Olivier Bodenreider
Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA
Exercise 1

- What are the Clinical Drug Components for Zyrtec? (RxNav)
Exercise 2

◆ What are the parts of the Aorta? (FMA)
Exercise 3

◆ What are the parents of Hodgkin’s disease in SNOMED CT?
  ● Try SNOMEDCTID: 118599009
◆ What is its associated morphology?
Exercise 4

◆ What are the various meanings of IL-2? (UMLS)
Exercise 5

- What are the pharmacologic actions of Zyrtec? (MeSH)
Exercise 6

- What are some synonyms for Schwannoma? (NCI Thesaurus)
Solutions
Exercise 1

◆ What are the Clinical Drug Components for Zyrtec? (RxNav)
Exercise 2

◆ What are the parts of the Aorta? (FMA)
Exercise 3

◆ What are the parents of Hodgkin’s disease in SNOMED DT?
  • Try SNOMEDCTID: 118599009
◆ What is its associated morphology?

<table>
<thead>
<tr>
<th>CONCEPT</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ID</td>
<td>Fully Specified Name</td>
<td>Concept Status</td>
<td>CTV3ID</td>
<td>SNOMED ID</td>
</tr>
<tr>
<td>118599009</td>
<td>Hodgkin's disease (disorder)</td>
<td>Current (0)</td>
<td>B61..</td>
<td>DC-F1000</td>
</tr>
</tbody>
</table>

| PARENTS | | | |
| --- | --- | --- | |
| Concept ID  | FSN for Parent Concept (This Concept IS A) | |
| 118600007 | Malignant lymphoma (disorder) | |

| ATTRIBUTES | | | |
| --- | --- | --- | |
| Concept ID  | FSN for Target Concept | Relationship Type | Values |
| 288526004 | Episodicities (qualifier value) | Episodicity (attribute) | (New episode; Ongoing episode; Old episode; Undefined episodicity; Other episode RCGP; First episode) |
Exercise 4

- What are the various meanings of IL-2? (UMLS)

![Metathesaurus Search]

- IL2 gene [C0879590]
- Interleukin-2 [C0021756]
- Recombinant Interleukin-2 [C1522405]
- interleukin-2 binding [C1149229]
Exercise 5

◦ What are the pharmacologic actions of Zyrtec? (MeSH)

| Pharm. Action | Anti-Allergic Agents |
| Pharm. Action | Histamine H1 Antagonists, Non-Sedating |
Exercise 6

What are some synonyms for Schwannoma? (NCI Thesaurus)

<table>
<thead>
<tr>
<th>Preferred_Name</th>
<th>Schwannoma</th>
</tr>
</thead>
<tbody>
<tr>
<td>Semantic_Type</td>
<td>Neoplastic Process</td>
</tr>
<tr>
<td>Synonym</td>
<td>Neurilemmoma</td>
</tr>
<tr>
<td>Synonym</td>
<td>Neurinoma</td>
</tr>
<tr>
<td>Synonym</td>
<td>Schwannoma</td>
</tr>
<tr>
<td>Synonym</td>
<td>schwannoma</td>
</tr>
</tbody>
</table>
What to look for

◆ Search modalities
  - Spelling correction / auto-completion / normalization
  - Word combinations (AND/OR)

◆ Visualization
  - Graph vs. (forest of) trees

◆ Navigation

◆ What properties are displayed
## Agenda

<table>
<thead>
<tr>
<th>Monday, June 9</th>
<th>Introduction to Biomedical Ontologies</th>
<th>Design Principles, Formalisms and Tools for Biomedical Ontologies</th>
<th>Biomedical Ontologies - Content and structure - Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td>Extending Biomedical Ontologies</td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>

Lister Hill National Center for Biomedical Communications
Contrasting and Critiquing Biomedical Ontologies

Olivier Bodenreider
Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Exercise #1

- Hodgkin’s disease
  - NCI Thesaurus
  - SNOMED CT
Exercise #2

- Prostate
  - FMA
  - SNOMED CT
Exercise #3

- Cetirizine
  - MeSH
  - SNOMED CT
Solutions
Solutions

*Exercise #1*
Exercise #1

◆ Hodgkin’s disease
  ● NCI Thesaurus
    ■ Using the NCI browser (EVS)
      http://nciterms.nci.nih.gov/
  ● SNOMED CT
    ■ Using the online browser from U. Sydney
Hodgkin’s disease in NCIt (1)

**Hodgkin Lymphoma**

<table>
<thead>
<tr>
<th>Identifiers:</th>
</tr>
</thead>
<tbody>
<tr>
<td>name</td>
</tr>
<tr>
<td>code</td>
</tr>
</tbody>
</table>

**Relationships to other concepts:**

- Disease_Has_Primary_Anatomic_Site
- Disease_Has_Normal_Tissue_Origin
- Disease_Excludes_Normal_Cell_Origin
- Disease_Excludes_Normal_Cell_Origin
- Disease_Has_Abnormal_Cell
- Disease_Has_Associated_Anatomic_Site
- Disease_Has_Normal_Cell_Origin
- Disease_Has_Primary_Anatomic_Site

Superconcepts:

- Common Hematopoietic Neoplasm
- Lymphoma

Lister Hill National Center for Biomedical Communications
## Hodgkin’s disease in NCIt (1)

### Information about this concept:

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALT_DEFINITION</td>
<td>NCI-GLOSS</td>
</tr>
<tr>
<td>DEFINITION</td>
<td>NCI</td>
</tr>
<tr>
<td>ICD-0-3_Id</td>
<td>9650/3</td>
</tr>
<tr>
<td>Preferred_Name</td>
<td>Hodgkin Lymphoma</td>
</tr>
<tr>
<td>Semantic_Type</td>
<td>Neoplastic Process</td>
</tr>
<tr>
<td>Synonym</td>
<td>HL</td>
</tr>
<tr>
<td>Synonym</td>
<td>Hodgkin Lymphoma</td>
</tr>
<tr>
<td>Synonym</td>
<td>Hodgkin's Disease</td>
</tr>
<tr>
<td>Synonym</td>
<td>Hodgkin's Lymphoma</td>
</tr>
<tr>
<td>Synonym</td>
<td>Hodgkin's disease</td>
</tr>
<tr>
<td>Unified Medical Language System Concept Identifier</td>
<td>C0019829</td>
</tr>
</tbody>
</table>
Comments on Hodgkin’s disease in NCIt (1)

◆ Search term: “Hodgkin’s disease”
  ● Not found, although “Hodgkin’s disease” is listed as a synonym
  ● Search on “hodgkin”, select “Hodgkin lymphoma”

◆ Parent classes
  ● Common hematopoietic neoplasm
    ■ Not an ontological category
    ■ Would be better represented through an associative relation (e.g., along the lines of “has_prevalence high prevalence”)
    ■ Isa overloading
Comments on Hodgkin’s disease in NCIt (2)

◆ Associative relations

- For cancers, anatomy and morphology are foundational relations
- Here
  - Anatomy: Disease_Has_Primary_Anatomic_Site
    Hematopoietic and Lymphatic System
  - Morphology: not directly represented
    (indirectly through Disease_Has_Normal_Cell_Origin Mature Lymphocyte)
Hodgkin’s disease in SNOMED CT (1)

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>Fully Specified Name</th>
<th>Concept Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>118599009</td>
<td>Hodgkin's disease (disorder)</td>
<td>Current (0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Parent Concept (This Concept)</th>
</tr>
</thead>
<tbody>
<tr>
<td>118600007</td>
<td>Malignant lymphoma (disorder)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Target Concept</th>
<th>Relationship Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>288526004</td>
<td>Episodicities (qualifier value)</td>
<td>Episodicity (attribute)</td>
</tr>
<tr>
<td>128930002</td>
<td>Hodgkin lymphoma - category (morphologic abnormality)</td>
<td>Associated morphology (attribute)</td>
</tr>
</tbody>
</table>
Hodgkin’s disease in SNOMED CT (2)

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>Fully Specified Name</th>
<th>Concept Status</th>
<th>CTV3ID</th>
</tr>
</thead>
<tbody>
<tr>
<td>118599009</td>
<td>Hodgkin’s disease (disorder)</td>
<td>Current (0)</td>
<td>B61..</td>
</tr>
</tbody>
</table>

**DESCRIPTIONS and SYNONYMS**

<table>
<thead>
<tr>
<th>Description ID</th>
<th>Term</th>
<th>Description Status</th>
<th>Description Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hodgkin’s disease (clinical)</td>
<td>Current (0)</td>
<td>Preferred (1)</td>
</tr>
<tr>
<td>2</td>
<td>Malignant Hodgkin’s lymphoma</td>
<td>Current (0)</td>
<td>Synonym (2)</td>
</tr>
<tr>
<td>3</td>
<td>HD - Hodgkin's disease</td>
<td>Current (0)</td>
<td>Synonym (2)</td>
</tr>
</tbody>
</table>
Comments on Hodgkin’s disease in SNOMED CT (1)

◆ Search term: “Hodgkin’s disease”
  ● Not found, although “Hodgkin’s disease” is listed as a synonym
    ■ Search result: “Hodgkin lymphoma, nodular sclerosis, grade 1 (morphologic abnormality)”
  ● Search on “lymphoma”, navigate down from “Malignant lymphoma”
  ● “hodgkin’s disease” is ambiguous
    ■ Hodgkin lymphoma, no ICD-O subtype (morphologic abnormality)
    ■ Hodgkin's disease (disorder)
  ● “Malignant lymphoma, Hodgkin's”
    ■ NB: lymphoma is always malignant

◆ Parent classes
  ● Malignant lymphoma (clinical) [OK]
Comments on Hodgkin’s disease in SNOMED CT (2)

◆ Associative relations

- For cancers, anatomy and morphology are foundational relations
- Here
  - Anatomy: not directly represented
    (indirectly through descendant concepts, e.g., Hodgkin's disease of intrathoracic lymph nodes)
  - Morphology: Associated morphology Hodgkin lymphoma - category
Hodgkin’s disease  NCIt vs. SNOMED CT (1)

◆ Shared synonyms: NCIt 1/2, SNOMED CT 1/3
  ● Hodgkin’s disease

◆ Shared relations
  ● Isa
    ■ NCIt: Lymphoma
      – Definition: “malignant (clonal) proliferation of B-lymphocytes or T-lymphocytes which involves the lymph nodes, bone marrow and/or extranodal sites. This category includes Non-Hodgkin lymphomas and Hodgkin lymphomas.”
    ■ SNOMED CT: Malignant lymphoma
    ■ Same UMLS concept (CUI: C0024299)
Hodgkin’s disease  NCIt vs. SNOMED CT (2)

◆ Shared relations: Associative relations
  • Anatomy
    ■ In NCIt, but not in SNOMED CT
  • Morphology
    ■ In SNOMED CT, but not in NCIt
      – Only indirectly, though cell type
  • Cell type
    ■ Only in NCIt
Solutions

Exercise #2
Exercise #2

◆ Prostate
  ● FMA
    ■ Using the Foundational Model Explorer
      http://sig.biostr.washington.edu/projects/fm/FME/
  ● SNOMED CT
    ■ Using the online browser from U. Sydney
Prostate in FMA (1)
Prostate in FMA (2)
Comments on Prostate in FMA

◆ No synonyms in English
  ● Latin and Spanish synonyms
◆ Hierarchies
  ● Isa: Lobular organ
  ● Part_of: Set of pelvic viscera
◆ Associative relations
  ● Lymphatic drainage
  ● No spatial relations
Prostate in SNOMED CT (1)

435 results found for prostate:

<table>
<thead>
<tr>
<th>#</th>
<th>Concept ID</th>
<th>Fully Specified Name</th>
<th>Preferred Terms and Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>9713002</td>
<td>Prostatitis (disorder)</td>
<td>Inflammation of prostate- Prostatitis [PT]- Prostatitis, NOS</td>
</tr>
<tr>
<td>2</td>
<td>11441004</td>
<td>Prostatism (disorder)</td>
<td>Prostatism [PT]- Prostatism, NOS</td>
</tr>
<tr>
<td>3</td>
<td>41216001</td>
<td>Prostatic structure (body structure)</td>
<td>Prostatic structure [PT]- Prostate- Prostate, NOS</td>
</tr>
<tr>
<td>4</td>
<td>181422007</td>
<td>Entire prostate (body structure)</td>
<td>Entire prostate [PT]- Prostate</td>
</tr>
</tbody>
</table>

CONCEPT

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>Fully Specified Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>181422007</td>
<td>Entire prostate (body structure)</td>
</tr>
</tbody>
</table>

DESCRIPTIONS and SYNONYMS

<table>
<thead>
<tr>
<th>Description ID</th>
<th>Term</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Entire prostate</td>
</tr>
<tr>
<td>2</td>
<td>Prostate</td>
</tr>
</tbody>
</table>

PARENTS

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Parent Concept (This Concept IS A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male internal genital organ (body structure)</td>
</tr>
<tr>
<td>2</td>
<td>Prostatic structure (body structure)</td>
</tr>
<tr>
<td>3</td>
<td>Entire male genital organ (body structure)</td>
</tr>
</tbody>
</table>
Prostate in SNOMED CT (2)

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Target Concept</th>
<th>Relationship Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Entire viscus (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>2</td>
<td>Entire urinary tract (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>3</td>
<td>Entire male internal genitalia (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>4</td>
<td>Entire lower male genitourinary tract (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>5</td>
<td>Entire minor pelvis (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>6</td>
<td>Entire lower genitourinary tract (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>7</td>
<td>Entire male genital system (body structure)</td>
<td>Part of (attribute)</td>
</tr>
<tr>
<td>8</td>
<td>Entire abdomen (body structure)</td>
<td>Part of (attribute)</td>
</tr>
</tbody>
</table>
Comments on Prostate in SNOMED CT

- “Ambiguous” term
  - Entire prostate
  - Prostatic structure

- Structure-Entire-Part representation of anatomical entities in SNOMED CT
  - Reification of part_of
  - Enables mereological inference through isa hierarchy
  - Not intuitive
Structure-Entire-Part (SEP) triples

- S – The entity or any of its parts
- E – The entire anatomical entity
- P – Any parts of the anatomical entity

[Schulz & al., 1997]
[Schulz & al., 1998]
[Bodenreider et al., 2006]
Mereological inference through *isa*

Diagram illustrating "isa" and "part_of" relationships:

- Kidney *structure* isisa Entire kidney
- Kidney *part* is part_of Entire kidney
- Structure of layer of kidney
- Structure of region of kidney
- Entire cortex of kidney
- Entire pole of kidney

Lister Hill National Center for Biomedical Communications
274
FMA mapping goes to *Entire*

- **FMA**
- **SNOMED CT**

![Diagram](image)

- Kidney
- Kidney structure
  - Syn: Kidney
- Kidney part
- Entire kidney
  - Syn: Kidney
  - part_of
Prostate  FMA vs. SNOMED CT

◆ Shared synonyms: FMA 1/1, SNOMED CT 1/2
  ● Prostate

◆ Shared relations
  ● Isa: no
    ■ FMA
      – Lobular organ
    ■ SNOMED CT
      – Prostatic structure
      – Male internal genital organ
      – Entire male genital organ
Prostate  FMA vs. SNOMED CT

◆ Shared relations

• *Part of*: almost
  ■ FMA
    – Genital system
    – Content of *male* pelvis
    – Set of male pelvic viscera
    – Set of pelvic viscera
  ■ SNOMED CT
    – Entire minor pelvis
    – Entire *male* genital system
    – ...

Lister Hill National Center for Biomedical Communications 277
Solutions

Exercise #3
Exercise #3

- Cetirizine
  - MeSH
    - Using the MeSH browser
  - SNOMED CT
    - Using the online browser from U. Sydney
Cetirizine in MeSH (1)

<table>
<thead>
<tr>
<th>Entry Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2-(4-((4-Chlorophenyl)phenylmethyl)-1-piperazinyl)ethoxy)acetic Acid</td>
<td></td>
</tr>
<tr>
<td>Alerisin</td>
<td></td>
</tr>
<tr>
<td>Aliud Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>Alpharma Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>AWD.pharma Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>Azupharma Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>Basics Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>Cetalerg</td>
<td></td>
</tr>
<tr>
<td>Voltric</td>
<td></td>
</tr>
<tr>
<td>Wolff Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>Wörwag Brand of Cetirizine Dihydrochloride</td>
<td></td>
</tr>
<tr>
<td>Zetir</td>
<td></td>
</tr>
<tr>
<td>Zirtek</td>
<td></td>
</tr>
<tr>
<td>Zyrtec</td>
<td></td>
</tr>
</tbody>
</table>
Cetirizine in MeSH (2)

<table>
<thead>
<tr>
<th>Pharm. Action</th>
<th>Anti-Allergic Agents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pharm. Action</td>
<td>Histamine H1 Antagonists, Non-Sedating</td>
</tr>
</tbody>
</table>
Comments on Cetirizine in MeSH

◆ 45 entry terms
  ● Various generic and brand names
  ● Chemical formula
  ● Code (P-071)

◆ Hierarchy
  ● Isa: Piperazines [chemistry]

◆ Pharmacologic action
  ● Anti-Allergic Agents
  ● Histamine H1 Antagonists, Non-Sedating
Cetirizine in SNOMED CT (1)

15 results found for cetirizine:

<table>
<thead>
<tr>
<th>#</th>
<th>Concept ID</th>
<th>Fully Specified Name</th>
<th>Preferred Terms and Synonyms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1086550000</td>
<td>Cetirizine (product)</td>
<td>Cetirizine [PT]</td>
</tr>
<tr>
<td>2</td>
<td>372523007</td>
<td>Cetirizine (substance)</td>
<td>Cetirizine [PT]</td>
</tr>
</tbody>
</table>

**Cetirizine (substance)**

<table>
<thead>
<tr>
<th>CONCEPT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ID</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>372523007</td>
</tr>
</tbody>
</table>

**DESCRIPTIONS and SYNONYMS**

<table>
<thead>
<tr>
<th>DESCRIPTIONS and SYNONYMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Description ID</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>1211057019</td>
</tr>
</tbody>
</table>

**PARENTS**

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Parent Concept (This Concept IS A)</th>
</tr>
</thead>
<tbody>
<tr>
<td>372624008</td>
<td>Non-sedating antihistamine (substance)</td>
</tr>
</tbody>
</table>

**CHILDREN**

<table>
<thead>
<tr>
<th>Concept ID</th>
<th>FSN for Child Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td>108656004</td>
<td>Cetirizine hydrochloride (substance)</td>
</tr>
</tbody>
</table>
Cetirizine in SNOMED CT (2)

<table>
<thead>
<tr>
<th>CONCEPT</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ID</td>
<td>Fully Specified Name</td>
<td>Concept Status</td>
<td>CTV3ID</td>
<td></td>
</tr>
<tr>
<td>108655000</td>
<td>Cetirizine (product)</td>
<td>Current (0)</td>
<td></td>
<td>01Dq</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DESCRIPTIONS and SYNONYMS</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Description ID</td>
<td>Term</td>
<td>Description Status</td>
<td>Description Type</td>
</tr>
<tr>
<td>1</td>
<td>173189012</td>
<td>Cetirizine</td>
<td>Current (0)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PARENTS</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ID</td>
<td>FSN for Parent Concept (This Concept IS A)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>340956006</td>
<td>Non-sedating antihistamine (product)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ATTRIBUTES</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ID</td>
<td>FSN for Target Concept</td>
<td>Relationship Type</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>372523007</td>
<td>Cetirizine (substance)</td>
<td>Has active ingredient (attribute)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHILDREN</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Concept ID</td>
<td>FSN for Child Concept</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>320818006</td>
<td>Cetirizine dihydrochloride 10mg tablet (product)</td>
</tr>
<tr>
<td>2</td>
<td>320820009</td>
<td>Cetirizine dihydrochloride 1mg/1mL s/f liquid (product)</td>
</tr>
<tr>
<td>3</td>
<td>371746005</td>
<td>Cetirizine dihydrochloride 5mg tablet (product)</td>
</tr>
<tr>
<td>4</td>
<td>375571002</td>
<td>Cetirizine hydrochloride 5mg tablet (product)</td>
</tr>
<tr>
<td>5</td>
<td>375572009</td>
<td>Cetirizine hydrochloride 10mg tablet (product)</td>
</tr>
<tr>
<td>6</td>
<td>375573004</td>
<td>Cetirizine hydrochloride 5mg/5 mL syrup (product)</td>
</tr>
<tr>
<td>7</td>
<td>400462001</td>
<td>Cetirizine hydrochloride+pseudoephedrine hydrochloride (product)</td>
</tr>
<tr>
<td>8</td>
<td>409491005</td>
<td>Cetirizine hydrochloride 5mg chewable tablet (product)</td>
</tr>
<tr>
<td>9</td>
<td>409492003</td>
<td>Cetirizine hydrochloride 10mg chewable tablet (product)</td>
</tr>
</tbody>
</table>
Comments on Cetirizine in SNOMED CT

- **Ambiguous term**
  - Cetirizine (product)
  - Cetirizine (substance)

- **Hierarchy**
  - *Isa*: Non-sedating antihistamine (substance) [pharmacologic action]

- **No associative relations**
Cetirizine MeSH vs. SNOMED CT (1)

- Shared synonyms: MeSH 1/45, SNOMED CT 1/1
  - Cetirizine

- Shared relations: none

- MeSH:
  - Isa: <chemistry>
  - Associative: <pharmacologic action>

- SNOMED CT
  - Isa: <pharmacologic action>
  - Associative: none
Cetirizine MeSH vs. SNOMED CT (2)
Summary

◆ Differing representations
  ● Not necessarily inconsistent
  ● Consistency may be difficult to assess automatically

◆ Often due to idiosyncratic representation in one ontology

◆ Hindrance to ontology alignment and evaluation methods relying on shared relations
## Agenda

<table>
<thead>
<tr>
<th>Monday, June 9</th>
<th>Introduction to Biomedical Ontologies</th>
<th>Design Principles, Formalisms and Tools for Biomedical Ontologies</th>
<th>Biomedical Ontologies - Content and structure - Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td>Extending Biomedical Ontologies</td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>
Extending Biomedical Ontologies

June 9, 2008 – Session #2

Olivier Bodenreider
Lister Hill National Center for Biomedical Communications
Bethesda, Maryland - USA
Overview

- Corpus terminology
- Identify terms in biomedical text (in reference to the UMLS)
- Identify additional terms
- Place these terms in UMLS hierarchies

[Bodenreider, ACL 2002]
Free cortisol in sepsis and septic shock.


Department of Intensive Care, Kuopio University Hospital, PL 16222 Kuopio, Finland. Stepani.Bendel@kuh.fi

BACKGROUND: Severe sepsis activates the hypothalamo-pituitary axis, increasing cortisol production. In some studies, hydrocortisone substitution based on an adrenocorticotropic hormone-stimulation test or baseline cortisol measurement has improved outcome. Because only the free fraction of cortisol is active, measurement of free cortisol may be more important than total cortisol in critically ill patients. We measured total and free cortisol in patients with severe sepsis and related the concentrations to outcome. METHODS: In a prospective study, severe sepsis was defined according the American College of Chest Physicians/Society of Critical Care Medicine criteria. Blood samples were drawn within 24 h of study entry. Serum cortisol was analyzed by electrochemiluminescence immunoassay. The Coolens method was used for calculating serum free cortisol concentrations. RESULTS: Blood samples were collected from 125 patients, of whom 62 had severe sepsis and 63 septic shock. Hospital mortality was 21%. Calculated free serum cortisol correlated well with serum total cortisol ($r = 0.90$, $P < 0.001$). There was no difference in the total cortisol concentrations in patients with sepsis and septic shock ($728 \pm 386$ nmol/L vs $793 \pm 439$ nmol/L, $P = 0.44$). Non-survivors had higher calculated serum free ($209 \pm 151$ nmol/L) and total ($980 \pm 458$ nmol/L) cortisol concentrations than survivors ($119 \pm 111$ nmol/L, $P = 0.002$, and $704 \pm 383$ nmol/L, $P = 0.002$). Depending on the definition, the incidence of adrenal insufficiency varied from 8% to 54%. CONCLUSIONS: Clinically, calculation of free cortisol does not provide essential information for identification of patients who would benefit from corticoid treatment in severe sepsis and septic shock.
Identify UMLS concepts with MetaMap

http://skr.nlm.nih.gov/

Lister Hill National Center for Biomedical Communications
Interactive mode

Please Note: Users are responsible for compliance with the UMLS copyright restrictions.

To use this application, you must have signed the UMLS agreement. The UMLS agreement requires those who use the UMLS to file a brief report once a year to summarize their use of the UMLS. It also requires the acknowledgment that the UMLS contains copyrighted material and that those copyright restrictions be respected. The UMLS agreement requires users to agree to obtain agreements for EACH copyrighted source prior to its use within a commercial or production application.

[ Use of all the sources is permitted if the application is used for research purposes only. ]
Interactive MetaMap

User: umls1  NLM » LHNBC » SKR » Interactive

Please NOTE:
The Interactive mode is only intended for the testing of the various programs and their options.
BACKGROUND: Severe sepsis activates the hypothalamopituitary axis, increasing cortisol production. In some studies, hydrocortisone substitution based on an adrenocorticotropic hormone-stimulation test or baseline cortisol measurement has improved outcome. Because only the free fraction of cortisol is active, measurement of free cortisol may be more important than total cortisol in critically ill patients. We measured total and free cortisol in patients with severe sepsis and related the concentrations to outcome. METHODS: In a prospective study, severe sepsis was defined according the American College of Chest Physicians/Society of Critical Care Medicine criteria. Blood samples were drawn within 24 h of study entry. Serum cortisol was analyzed by electrochemiluminescence immunoassay. The Coolens method was used for
Select options

<table>
<thead>
<tr>
<th>Output Display Options</th>
<th>Output Display Options (continued)</th>
<th>Behavior Options (continued)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tagger Output (-T)</td>
<td>Show Original Phrases (-H)</td>
<td>Prefer Multiple Concepts (-Y)</td>
</tr>
<tr>
<td>Variants (-v)</td>
<td>Show Concept's Sources (-G)</td>
<td>Best Mappings Only (-b)</td>
</tr>
<tr>
<td>Plain Syntax (-p)</td>
<td>Show Acronym/Abbreviations (-j)</td>
<td>Truncate Candidates Mapping (-X)</td>
</tr>
<tr>
<td>Syntax (-x)</td>
<td></td>
<td>Use Word Sense Disambiguation (-y)</td>
</tr>
<tr>
<td>Candidates (-c)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number Candidates (-n)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Semantic Types (-s)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Show CUIs (-l)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mappings (-m)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Show Preferred Names Only (-O)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMI Output (-M)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Machine Output (-q)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fielded Output (-f)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Formal Tagger Output (-F)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fielded MMI output (-N)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Behavior Options      |                                 |                             |
|-----------------------|                                 |                             |
| Tag Text (-t)          |                                 |                             |
| No Derivational Variants (-d) |           |                             |
| Adj/Noun Derivational Variants (-D) |       |                             |
| No Acronym/AmbigVation Variants (-a) |       |                             |
| Unique Acronym/Abbrev Vairs (-u) |              |                             |
| Ignore Stop Phrases (-K) (System Use) |         |                             |
| Stop Large N (-I)     |                                 |                             |
| Threshold (-t):        |                                 |                             |
| Ignore Word Order (-i) |                                 |                             |

| Browse Mode Options   |                                 |                             |
|-----------------------|                                 |                             |
| Term Processing (-z)  |                                 |                             |
| Allow Overmatches (-o) |                                 |                             |
| Allow Concept Gaps (-g) |                          |                             |

| Misc. Options         |                                 |                             |
|-----------------------|                                 |                             |
| Display Phrases Only  |                                 |                             |
| Dynamic Variant Generation (-g) |                    |                             |
Run MetaMap

<table>
<thead>
<tr>
<th>&quot;Restrict to&quot; or &quot;Exclude&quot; Vocabulary Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Restrict to Sources (-R)</td>
</tr>
<tr>
<td>□ Exclude Sources (-e)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>&quot;Restrict to&quot; or &quot;Exclude&quot; Semantic Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>□ Restrict to Semantic Type(s) (-J)</td>
</tr>
<tr>
<td>□ Exclude Semantic Type(s) (-k)</td>
</tr>
</tbody>
</table>

Submit Interactive MetaMap  Reset Form
Processing 00000000.tx.1: BACKGROUND: Severe sepsis activates the hypothalamic-pituitary axis, increasing cortisol production.

Phrase: "Severe sepsis"
>>>>> Phrase
severe sepsis
<<<< Phrase
>>>>> Candidates
Meta Candidates (8):
  1000 C1719672:Severe Sepsis [Disease or Syndrome]
  861 C0036690:Sepsis (Septicemia) [Disease or Syndrome]
  861 C0243026:Sepsis (Systemic infection) [Disease or Syndrome]
  861 C1090821:Sepsis [Invertebrate]
  789 C0333534:Septic [Functional Concept]
  694 C0205082:Severe [Qualitative Concept]
  694 C1519275:SEVERE (Severe Adverse Event) [Finding]
  694 C1561581:Severe (Allergy Severity - Severe) [Finding]
<<<< Candidates
>>>>> Mappings
Meta Mapping (1000):
  1000 C1719672:Severe Sepsis [Disease or Syndrome]
<<<< Mappings
Suggest term candidates

- Not recognized by MetaMap at all
- Partially identified by MetaMap
- Missing terms in a concept
Suggest placement in UMLS

- Use a browser
- Identify close parent
- Examine its children
- Assess placement by comparing with potential siblings
Possible new terms (1)

◆ Hypothalamopituitary axis
  ● Concept exists: C0678897, but missing exact (neoclassical) synonym
    ■ hypothalamic pituitary axis
    ■ hypothalamus hypophysis axis
    ■ hypothalamus-pituitary axis

◆ American College of Chest Physicians
  ● Similar to other American Colleges (e.g., American College of Physicians ()
  ● Integrate as a child of Professional Organization or Group (C1522486 )
  ● NB: instance, cannot be a child of ACP
Possible new terms (2)

◆ Free cortisol
  ● Identified as a substance (C0443476), not a laboratory procedure / test result
    ■ Cortisol, free measurement (C0236401)

◆ Coolens method
  ● Missing term / concept
  ● Method for estimating (not measuring directly) the free fraction of cortisol
Possible new terms (3)

◆ Electrochemiluminescence immunoassay
  ● Missing concept
  ● Create as a child of Chemiluminescence assay (C0201709)

◆ Nonsurvivors
  ● Survivors exists as a concept (C0206194)
  ● Create as a child of Patients (C0030705)
# Agenda

<table>
<thead>
<tr>
<th>Date</th>
<th>Session 1</th>
<th>Session 2</th>
<th>Session 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monday, June 9</td>
<td>Introduction to Biomedical Ontologies</td>
<td>Design Principles, Formalisms and Tools for</td>
<td>Biomedical Ontologies - Content and structure</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Biomedical Ontologies</td>
<td>- Function</td>
</tr>
<tr>
<td>Tuesday, June 10</td>
<td>Interfaces to Biomedical Ontologies</td>
<td>Searching and Analyzing Biomedical Concepts</td>
<td>Contrasting Biomedical Ontologies</td>
</tr>
<tr>
<td>Wednesday, June 11</td>
<td>Critical Analysis of Biomedical Ontologies</td>
<td>Extending Biomedical Ontologies</td>
<td>Using Biomedical Ontologies for Data Integration</td>
</tr>
</tbody>
</table>
Using Biomedical Ontologies
for Data Integration

Olivier Bodenreider
Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA
Overview

◆ Motivation
◆ Some practical considerations and issues
  ● Integration approaches
  ● Concept repositories
  ● Using existing mappings
  ● Creating mappings through the UMLS
  ● Comparing semantic descriptions
◆ Thinking outside the integration box
Motivation
Motivation  Translational research

◆ “Bench to Bedside”

◆ Integration of clinical and research activities and results

◆ Supported by research programs
  ● NIH Roadmap
  ● Clinical and Translational Science Awards (CTSA)

◆ Requires the effective integration and exchange and of information between
  ● Basic research
  ● Clinical research
TRANSLATIONAL RESEARCH

OVERVIEW

To improve human health, scientific discoveries must be translated into practical applications. Such discoveries typically begin at the "bench" with basic research — in which scientists study disease at a molecular or cellular level — then progress to the clinical level, or the patient’s "bedside."

Scientists are increasingly aware that this bench-to-bedside approach to translational research is really a two-way street. Basic scientists provide clinicians with new tools for use in patients and for assessment of their impact, and clinical researchers make novel observations about the nature and progression of disease that often stimulate basic investigations.
Motivation  Translational research

Basic Research

Clinical Research and Practice
Terminology and translational research

Cancer Basic Research

NCI Thesaurus

SNOMED CT

EHR Cancer Patients
Some practical considerations and issues

Integration approaches
Approaches to data integration

◆ Warehousing
  - Sources to be integrated are transformed into a common format and converted to a common vocabulary
  - Normalization through ontologies (e.g., GO annotations)

◆ Mediation
  - Local schema (of the sources)
  - Global schema (in reference to which the queries are made)
  - Ontologies help define the global schema and map between local and global schemas (OntoFusion, ARIANE)
Some practical considerations and issues

Concept repositories
(Integrated) concept repositories

- Unified Medical Language System
- NCBO’s BioPortal
  http://www.bioontology.org/tools/portal/bioportal.html
- Open Biomedical Ontologies (OBO)
  http://obofoundry.org/
- caDSR
  http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/cadsr
Integrating subdomains

Clinical repositories

Genetic knowledge bases

Other subdomains

SNOMED CT

OMIM

UMLS

MeSH

Biomedical literature

NCBI Taxonomy

Model organisms

Anatomy

GO

FMA

Genome annotations

Lister Hill National Center for Biomedical Communications
Trans-namespave integration

Addison's disease (363732003)

Other subdomains

Clinical repositories

Snomed CT

Genetic knowledge bases

OMIM

Addison Disease (D000224)

SNOMED CT

Lister Hill National Center for Biomedical Communications

Anatomy

Genome annotations

UMLS

Biomedical literature

NCBI Taxonomy

GO

Model organisms

FMA

Genome annotations

Lister Hill National Center for Biomedical Communications

MeSH

C0001403

Biomedical literature
Some practical considerations and issues

Mappings
Mappings

UMLS

NCI Thesaurus

SNOMED CT
Mappings

- Created manually
  - UMLS
- Created automatically
  - BioPortal

- Key to enabling semantic interoperability
- Enabling resource for the Semantic Web
Quality of mappings

◆ Created for a purpose
  ● Reusability issues

◆ Generally unidirectional
  ● Mapping from ontology 1 to ontology 2
  ● Not necessarily reversible
Some practical considerations and issues

Comparing semantic descriptions
SNOMED CT/NCI Thesaurus
Limited consistency

Lister Hill National Center for Biomedical Communications
Comparing formal definitions

◆ Relatively small proportion of relata in common between equivalent concepts from NCIt and SNOMED CT

◆ Large number of primitive concepts in NCIt and SNOMED CT (70-80%)

◆ Insufficient for effectively comparing definitions
  • Could not be used for validating the mapping provided by the UMLS

[Bodenreider, KRMED 2008]

Lister Hill National Center for Biomedical Communications 326
Exercises
Exercise #1

◆ Check the equivalence (shared relata) between these 2 concepts:
  ● NCI Thesaurus: N:Endocrine pancreas disorder
  ● SNOMED CT: S:Disorder of endocrine pancreas
Exercise #2

- Find a correspondence in SNOMED CT for the LOINC term: Sodium:SCnc:-Pt:Ser/Plas:Qn
  [the molar concentration of sodium is measured in the plasma (or serum), with quantitative result]

<table>
<thead>
<tr>
<th>Axis</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component</td>
<td>Sodium</td>
</tr>
<tr>
<td>Property</td>
<td>SCnc – Substance Concentration (per volume)</td>
</tr>
<tr>
<td>Timing</td>
<td>Pt – Point in time (Random)</td>
</tr>
<tr>
<td>System</td>
<td>Ser/Plas – Serum or Plasma</td>
</tr>
<tr>
<td>Scale</td>
<td>Qn – Quantitative</td>
</tr>
<tr>
<td>Method</td>
<td>--</td>
</tr>
</tbody>
</table>
Comments on exercise #2

- Difficult in the absence of a search mechanism on the values of the relations
- Large number of underspecified descriptions in SNOMED CT
- 2 separate concepts for plasma and serum concentrations of sodium in SNOMED CT
- Property, time and scale not represented in SNOMED CT
Thinking outside the integration box

The Butte approach
Integrating genomic and clinical data

Genomic data

Clinical data

Upregulated genes

Diseases (extracted from text)

Coded discharge summaries

Laboratory data

Lister Hill National Center for Biomedical Communications
References


Medical
Ontology
Research

Contact:  olivier@nlm.nih.gov
Web:  mor.nlm.nih.gov

Olivier Bodenreider
Lister Hill National Center
for Biomedical Communications
Bethesda, Maryland - USA