What Do Biomedical Terminologies Tell Us About Biomedicine?

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

◆ Why biomedical terminologies?
◆ Introduction to biomedical terminologies through an example
◆ Biomedical terms as names for biomedical classes
◆ Terminological relations as a surrogate for ontological relations
Why biomedical terminologies?
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 general Bill for this present year, ending the 19th of December 1664, according to the Report made to the KINGS most Excellent Majesty.

By the Company of Parish-Clerks of London, 

The Diseases and Casualties this year.

1. Smallpox 21
2. Plague 18
3. Pox 15
4. Plague 12
5. Pox 11
6. Plague 10
7. Pox 8
8. Plague 7
9. Pox 6
10. Plague 5

Lister Hill National Center for Biomedical Communications
Limitations of existing classifications

“There 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.
Mouse Genome Database and GO

Entrez Gene

Gene: NF2  **neurofibromatosis 2**  [*Mus musculus*]
GeneID: 18016  Locus tag: **MGI:97307**

### General gene information

#### Gene Ontology

**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
- IEA
- IMP  PubMed
- IEA
- IMP  PubMed
- 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

- Anatomy
  - 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)
Introduction to biomedical terminologies through an example
Guy’s Hospital, London
Thomas Addison (1795-1860)
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)
◆ …
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

- eponym
- symptoms
- clinical variants

Different terms:

- Addisonian syndrome
- Bronzed disease
- Addison melanoderma
- Asthenia pigmentosa
- Primary adrenal deficiency
- Primary adrenal insufficiency
- Primary adrenocortical insufficiency
- Chronic adrenocortical insufficiency
CHAPTER 4

Endocrine, nutritional and metabolic diseases (E00-E90)

Disorders of other endocrine glands (E20-E35)

E27 Other disorders of adrenal gland

E27.0 Other adrenocortical overactivity
Overproduction of ACTH, not associated with Cushing’s disease
Premature adrenarche
Excludes1: Cushing’s syndrome (E24.-)

E27.1 Primary adrenocortical insufficiency

Addison’s disease
Adrenocortical insufficiency NOS
Autoimmune adrenalitis
Excludes1: Addison only phenotype adrenoleukodystrophy (E71.428)
amyloidosis (E85)
tuberculous Addison’s disease (A18.7)
Waterhouse-Friderichsen syndrome (A39.1)

E27.2 Addisonian crisis
Adrenal crisis
Adrenocortical crisis

E27.3 Drug-induced adrenocortical insufficiency
Code first (T36-T50) to identify drug

E27.4 Other and unspecified adrenocortical insufficiency

Draft ICD-10-CM Tabular Page 180 June 2003
<table>
<thead>
<tr>
<th>Medical Subject Headings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MeSH Tree Structures</strong></td>
</tr>
</tbody>
</table>

### 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]
      - Autoimmune Diseases of the Nervous System [C20.111.258]
SNOMED CT

Hierarchy

- adrenal cortical hypofunction
  - Addison's disease
  - Addison's disease due to autoimmunity
  - Addison's disease with adrenoleucodystrophy
  - polyglandular autoimmune syndrome, type 1
  - tuberculous Addison's disease

Detail

Concept Status: Current

- Primitive
  - Is a
    - adrenal cortical hypofunction
  - Finding site
    - adrenal cortex structure
  - Qualifiers
    - Onset
      - sudden onset
      - gradual onset
    - Severity
      - severities
    - Episodicity
      - Episodicities
    - Course
      - courses

Legacy codes

- SNOMED: DB-70620
- 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
  - Common cold
  - Cold temperature
  - COLD
  - Chronic Obstructive Airway Disease

- **Truncated terms**
  - Calcium
  - Ca++
  - Coagulation factor IV
  - 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
  - …
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

Non-physical anatomical entity

Material physical anatomical entity

Non-material physical anatomical entity

Anatomical structure

Body substance

Anat. space

Anat. surface

Anat. line

Anat. point

Spatial dimension

Mass

3D 2D 1D 0D

Inherent 3D shape
No explicit classificatory principle
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

- Is a
  - Viral infections of the central nervous system
    - Infective meningitis
      - Causative agent
        - Virus
      - Group
        - Associated morphology
          - Inflammation
        - Finding site
          - Meninges structure
Underspecification of the relations

Diseases
  ↓
CNS diseases
  ↓
CNS infections
    ↓
Meningitis
    ↓
Viral meningitis
  ↓
Virus diseases
  ↓
CNS viral diseases

isa ?

parent

child
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
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