Controlled Vocabularies & Semantic Standards

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

- Explain how clinical features are reflected in disease names
- List and describe the main biomedical ontologies used in 21st century healthcare
- Discuss the purpose of biomedical ontologies in knowledge management, clinical decision support and analytics
- List and describe the main biomedical semantic standards
References

Additional references


Outline

◆ Introduction to biomedical terminologies through an example

◆ “High-Impact” Biomedical Ontologies
  ● Structural perspective

◆ Biomedical Ontologies “in Action”
  ● Functional perspective
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)
- ...


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

Synonyms:
- eponym
- symptoms
- clinical
- variants
Internal Classification of Diseases

IV Endocrine, nutritional and metabolic diseases

E00-E07 Disorders of thyroid gland
E10-E14 Diabetes mellitus
E15-E16 Other disorders of glucose regulation and pancreatic internal secretion

E20-E35 Disorders of other endocrine glands

E20 Hypoparathyroidism
E21 Hyperparathyroidism and other disorders of parathyroid gland
E22 Hyperfunction of pituitary gland
E23 Hypofunction and other disorders of pituitary gland

E24 Cushing syndrome
E25 Adrenogenital disorders
E26 Hyperaldosteronism

E27 Other disorders of adrenal gland

E27.0 Other adrenocortical overactivity
E27.1 Primary adrenocortical insufficiency

E27.2 Addisonian crisis

E27.3 Drug-induced adrenocortical insufficiency

E27.4 Other and unspecified adrenocortical insufficiency

E28 Ovarian dysfunction
E29 Testicular dysfunction
E30 Disorders of puberty, not elsewhere classified
E31 Pol glandular dysfunction
E32 Diseases of thymus
E34 Other endocrine disorders
E35 Disorders of endocrine glands in diseases classified elsewhere

E27.0 Other adrenocortical overactivity
Overproduction of ACTH, not associated with Cushing disease
Premature adrenarche

Excl.: Cushing syndrome (E24-)

E27.1 Primary adrenocortical insufficiency
Addison disease
Autoimmune adrenalitis

Excl.: amyloidosis (E85-)
tuberculous Addison disease (A16.7)
Waterhouse-Friderichsen syndrome (A39.1)

E27.2 Addisonian crisis
Adrenal crisis
Adrenocortical crisis

E27.3 Drug-induced adrenocortical insufficiency
Use additional external cause code (Chapter XX), if desired, to identify drug.

E27.4 Other and unspecified adrenocortical insufficiency
Adrenal:
• haemorrhage
• infarction
Adrenocortical insufficiency NOS
Hyperaldosteronism

Excl.: adrenoleukodystrophy [Addison-Schilder] (E71.3)
Waterhouse-Friderichsen syndrome (A39.1)

E27.5 Adrenomedullary hyperfunction
Adrenomedullary hyperplasia
Catecholamine hypersecretion

E27.8 Other specified disorders of adrenal gland
Abnormality of cortisol-binding globulin

E27.9 Disorder of adrenal gland, unspecified
Medical Subject Headings

MeSH Tree Structures

Endocrine System Diseases [C19]
- Adrenal Gland Diseases [C19.053]
  - Adrenal Insufficiency [C19.053.500]
    - Addison Disease [C19.053.500.263]
    - Adrenoleukodystrophy [C19.053.500.270]
    - Hypoaldosteronism [C19.053.500.480]
    - Waterhouse-Friderichsen Syndrome [C19.053.500.740]

Immune System Diseases [C20]
- Autoimmune Diseases [C20.111]
  - Addison Disease [C20.111.163]
    - Anemia, Hemolytic, Autoimmune [C20.111.175]
    - Anti-Glomerular Basement Membrane Disease [C20.111.190]
    - Anti-Neutrophil Cytoplasmic Antibody-Associated Vasculitis [C20.111.193] + Antiphospholipid Syndrome [C20.111.197]
    - Arthritis, Juvenile [C20.111.198]
    - Arthritis, Rheumatoid [C20.111.199] +
    - Autoimmune Diseases of the Nervous System [C20.111.258] + [...]

Lister Hill National Center for Biomedical Communications
Controlled Vocabularies & Semantic Standards
Part 2

“High-Impact” Biomedical Ontologies

A Structural Perspective
Overview

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

◆ “High-impact” biomedical ontologies [J. Cimino, YBMI 2006]
  ● 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)
International Classification of Diseases
ICD Characteristics (1)

- Current version: ICD-10 (2017)
  - Annual updates
- Type: Classification
- Domain: Disorders
- Developer: World Health Organization (WHO)
- Funding: WHO
- Publicly available: Yes
- Used for: Mortality and morbidity statistics worldwide
- URL: http://www.who.int/classifications/icd/en/
ICD Characteristics (2)

◆ Number of
  ● Concepts: 12,320 (ICD-10, 2004)
  ● 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)

◆ Specific coding rules

◆ Distribution: Proprietary format
ICD Top level

ICD-10 Version:2016

I Certain infectious and parasitic diseases
II Neoplasms
III Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
IV Endocrine, nutritional and metabolic diseases
V Mental and behavioural disorders
VI Diseases of the nervous system
VII Diseases of the eye and adnexa
VIII Diseases of the ear and mastoid process
IX Diseases of the circulatory system
X Diseases of the respiratory system
XI Diseases of the digestive system
XII Diseases of the skin and subcutaneous tissue
XIII Diseases of the musculoskeletal system and connective tissue
XIV Diseases of the genitourinary system
XV Pregnancy, childbirth and the puerperium
XVI Certain conditions originating in the perinatal period
XVII Congenital malformations, deformations and chromosomal abnormalities
XVIII Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
XIX Injury, poisoning and certain other consequences of external causes
XX External causes of morbidity and mortality
XXI Factors influencing health status and contact with health services
XXII Codes for special purposes
ICD Example

Idiosyncratic inclusion/exclusion criteria

Type 1 diabetes mellitus

Incl.: diabetes (mellitus):
- brittle
- juvenile-onset
- ketosis-prone

Excl.: 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)

E84 Cystic fibrosis

Incl.: mucoviscidosis

E84.0 Cystic fibrosis with pulmonary manifestations

E84.1 Cystic fibrosis with intestinal manifestations
  Distal intestinal obstruction syndrome
  Meconium ileus in cystic fibrosis† (P75*)

Excl.: meconium obstruction (ileus) in cases where cystic fibrosis is known not to be present (P76.0)

E84.8 Cystic fibrosis with other manifestations
E84.9 Cystic fibrosis, unspecified
**ICD-10-CM**

- Derived from: ICD-10
  - Finer-grained (both clinically and administratively)
- Type: Classification
  - 92,042 codes (2015)
  - Terms: 1.2 per concept
- Domain: Disorders
- Developer: National Center for Health Statistics (CDC/NCHS)
- Funding: U.S. Government
- Publicly available: Yes
- Used for: Billing
- URL: [http://www.cdc.gov/nchs/icd/icd10cm.htm](http://www.cdc.gov/nchs/icd/icd10cm.htm)
# ICD-10 vs. ICD-10-CM

## W58: Contact with crocodile or alligator

The appropriate 7th character is to be added to each code from category W58:
- A - initial encounter
- D - subsequent encounter
- S - sequela

### W58.0: Contact with alligator
- **W58.01** Bitten by alligator
- **W58.02** Struck by alligator
- **W58.03** Crushed by alligator
- **W58.09** Other contact with alligator

### W58.1: Contact with crocodile
- **W58.11** Bitten by crocodile
- **W58.12** Struck by crocodile
- **W58.13** Crushed by crocodile
- **W58.19** Other contact with crocodile

- **W58.01A** Bitten by alligator, initial encounter
- **W58.01D** Bitten by alligator, subsequent encounter
- **W58.01S** Bitten by alligator, sequela
Logical Observation Identifiers, Names and Codes (LOINC)
LOINC Characteristics (1)

- Current version: 2.61 (June 2017)
  - 2 annual releases
- Type: Controlled terminology*
- Domain: Laboratory and clinical observations
- Developer: Regenstrief Institute
- Funding: NLM and other sources
- Publicly available: Yes
- Used for: information exchange
- URL: [https://loinc.org/](https://loinc.org/)
LOINC Characteristics (2)

- Number of
  - Concepts: 73,958 active codes (2.52, June 2015)
  - Terms: 1 per concept (“long name”)

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

- Distribution: proprietary database format
**LOINC Example**

- Sodium [Moles/volume] in Serum or Plasma
  - [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>
2951-2  Sodium [Moles/volume] in Serum or Plasma

**NAME**

<table>
<thead>
<tr>
<th>Fully-Specified Name:</th>
<th>Component</th>
<th>Property</th>
<th>Time</th>
<th>System</th>
<th>Scale</th>
<th>Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sodium</td>
<td></td>
<td>SCnc</td>
<td>Pt</td>
<td>Ser/Plas</td>
<td>Qn</td>
<td></td>
</tr>
</tbody>
</table>

**PART DEFINITION/DESCRIPTION(S)**

Sodium is an essential nutrient that regulates blood volume, blood pressure, osmotic equilibrium and electrolyte balance. Sodium chloride is the principal source of sodium in the diet, and is used for seasoning and as a preservative. Increased levels of sodium intake can cause hypertension and reportedly leads to 7.6 million premature deaths worldwide. Sodium is also important in neuron function and osmoregulation between cells and the extracellular fluid.

**Copyright:** Text is available under the Creative Commons Attribution/Share-Alike License. See http://creativecommons.org/licenses/by-sa/3.0/ for details.


**BASIC ATTRIBUTES**

- **Class/Type:** CHEM/Lab
- **CDISC Lab Test:** Y
- **Common Lab Results Rank:** #5
- **Common SI Lab Results Rank:** #5
- **Common Orders Rank:** #107
- **Last Updated in Version:** 2.34
- **Order vs. Obs.:** Both
- **Status:** Active

**EXAMPLE UNITS**

<table>
<thead>
<tr>
<th>Unit</th>
<th>Source Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>mmol/L</td>
<td>EXAMPLE UCUM UNITS</td>
</tr>
<tr>
<td>mmol/L</td>
<td>REGENSTRIEF</td>
</tr>
<tr>
<td>mmol/L</td>
<td>eCHN</td>
</tr>
</tbody>
</table>

**UNITS AND RANGE**

<table>
<thead>
<tr>
<th>Range</th>
<th>Units Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>mmol/L [136,145]</td>
<td></td>
</tr>
</tbody>
</table>
SNOMED Clinical Terms
SNOMED CT Characteristics (1)

- Current version: July 31, 2017
  - 2 annual releases
- Type: Reference terminology / ontology
- Domain: Clinical medicine
- Developer: IHTSDO
- Funding: IHTSDO member countries
- Publicly available: Yes*
- Used for: clinical documentation, information exchange, analytics
- URL: http://www.ihtsdo.org/
SNOMED CT Characteristics (2)

- **Number of**
  - Concepts: 320,912 active concepts (Sept. 2016)
  - Terms: 2.6 per concept (“descriptions”)

- **Major organizing principles:**
  - Polyhierarchy
  - Rich set of associative relationships
  - Logical definitions (incomplete: many primitives)
  - Built using description logics (EL++)

- **Distribution:** RF2 (proprietary)
SNOMED CT Top level

- SNOMED CT Concept
  - Body structure (body structure)
  - Clinical finding (finding)
  - Environment or geographical location (environment / location)
  - Event (event)
  - Observable entity (observable entity)
  - Organism (organism)
  - Pharmaceutical / biologic product (product)
  - Physical force (physical force)
  - Physical object (physical object)
  - Procedure (procedure)
  - Qualifier value (qualifier value)
  - Record artifact (record artifact)
  - Situation with explicit context (situation)
  - SNOMED CT Model Component (metadata)
  - Social context (social concept)
  - Special concept (special concept)
  - Specimen (specimen)
  - Staging and scales (staging scale)
  - Substance (substance)
SNOMED CT Example

Parents
- Operation on appendix (procedure)
- Partial excision of large intestine (procedure)

Appendectomy (procedure)
SCTID: 80146002
80146002 | Appendectomy (procedure) |
  Appendectomy
  Excision of appendix
  Appendicectomy
  Appendectomy (procedure)

Procedure site - Direct → Appendix structure
Method → Excision - action

Children (8)
- Appendectomy with drainage (procedure)
- Emergency appendectomy (procedure)
- Excision of appendiceal stump (procedure)
- Excision of ruptured appendix by open approach (procedure)
- Incidental appendectomy (procedure)
- Interval appendectomy (procedure)
- Laparoscopic appendectomy (procedure)
- Non-emergency appendectomy (procedure)
SNOMED CT Example

80146002
Appendectomy (procedure)

27010001
Partial excision of large intestine (procedure)

8613002
Operation on appendix (procedure)

405813007
Procedure site - Direct (attribute)

66754008
Appendix structure (body structure)

260686004
Method (attribute)

129304002
Excision - action (qualifier value)
RxNorm  Characteristics (1)

- Current version: August 2017
  - Monthly releases (+weekly updates)
- Type: Controlled terminology
- Domain: Drug names
- Developer: NLM
- Funding: NLM
- Publicly available: Yes*
- Used for: e-prescribing, information exchange, analytics
- URL: http://www.nlm.nih.gov/research/umls/rxnorm/
**RxNorm Characteristics (2)**

- **Number of**
  - Concepts: 117,774 (March 2016)
  - Terms: 1.5 per concept

- **Major organizing principles:**
  - Generic vs. brand
  - Ingredient + Strength + Dose form
  - No hierarchical structure; rich graph of associative relations
  - Integrates all major US drug information sources
  - No clinical information

- **Distribution:** similar to UMLS RRF format
**RxNorm** Normalized form

<table>
<thead>
<tr>
<th>Strength</th>
<th>Ingredient</th>
<th>Dose form</th>
</tr>
</thead>
<tbody>
<tr>
<td>4mg/ml</td>
<td>Fluoxetine</td>
<td>Oral Solution</td>
</tr>
</tbody>
</table>

Semantic clinical drug component

Semantic clinical drug form

Semantic clinical drug
RxNorm Example

- **Ingredient**: Azithromycin

- **Brand Name**: Zithromax

- **C. Drug Comp.**: Azithromycin 250 MG
- **C. Drug Form**: Azithromycin Oral Tablet

- **B. Drug Comp.**: Azithromycin 250 MG
  
- **B. Drug Form**: Azithromycin Oral Tablet [Zithromax]

- **C. Drug**: Azithromycin 250 MG Oral Tablet

- **B. Drug**: Zithromax 250 MG Oral Tablet

- **G. Pack**: {6 (Azithromycin 250 MG Oral Tablet) } Pack

- **B. Pack**: Z-PAK
**Warfarin** [RxCUI = 11289]

<table>
<thead>
<tr>
<th>IN/MIN</th>
<th>Ingredient (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Warfarin</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PIN</th>
<th>Precise Ingredient (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Warfarin Potassium</td>
</tr>
<tr>
<td></td>
<td>Warfarin Sodium</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>BN</th>
<th>Brand Name (2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coumadin</td>
</tr>
<tr>
<td></td>
<td>Jantoven</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SCDC</th>
<th>Clinical Drug Component (11)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Warfarin Sodium 0.5 MG</td>
</tr>
<tr>
<td></td>
<td>Warfarin Sodium 1 MG</td>
</tr>
<tr>
<td></td>
<td>Warfarin Sodium 10 MG</td>
</tr>
<tr>
<td></td>
<td>Warfarin Sodium 2 MG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SBDC</th>
<th>Branded Drug Component (18)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Warfarin Sodium 1 MG [Coumadin]</td>
</tr>
<tr>
<td></td>
<td>Warfarin Sodium 10 MG [Coumadin]</td>
</tr>
<tr>
<td></td>
<td>Warfarin Sodium 10 MG [Jantoven]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SCD/GPCK</th>
<th>Clinical Drug or Pack (11)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Warfarin Sodium 0.5 MG Oral Tablet</td>
</tr>
<tr>
<td></td>
<td>Warfarin Sodium 1 MG Oral Tablet</td>
</tr>
<tr>
<td></td>
<td>Warfarin Sodium 10 MG Oral Tablet</td>
</tr>
<tr>
<td></td>
<td>Warfarin Sodium 2 MG Oral Tablet</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SBD/BPCK</th>
<th>Branded Drug or Pack (18)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coumadin 1 MG Oral Tablet</td>
</tr>
<tr>
<td></td>
<td>Coumadin 10 MG Oral Tablet</td>
</tr>
<tr>
<td></td>
<td>Coumadin 2 MG Oral Tablet</td>
</tr>
<tr>
<td></td>
<td>Coumadin 2.5 MG Oral Tablet</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SCDG</th>
<th>Clinical Dose Form Group (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Warfarin Injectable Product</td>
</tr>
<tr>
<td></td>
<td>Warfarin Oral Product</td>
</tr>
<tr>
<td></td>
<td>Warfarin Pill</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DFG</th>
<th>Dose Form Group (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Injectable Product</td>
</tr>
<tr>
<td></td>
<td>Oral Product</td>
</tr>
<tr>
<td></td>
<td>Pill</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SBDG</th>
<th>Branded Dose Form Group (4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coumadin Oral Product</td>
</tr>
<tr>
<td></td>
<td>Coumadin Pill</td>
</tr>
<tr>
<td></td>
<td>Jantoven Oral Product</td>
</tr>
<tr>
<td></td>
<td>Jantoven Pill</td>
</tr>
</tbody>
</table>
Controlled Vocabularies & Semantic Standards
Part 3

Biomedical Ontologies “in Action”

A Functional Perspective
Overview

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

◆ “High-impact” biomedical ontologies

◆ 3 major categories of use
  ● Knowledge management
    ◦ Annotating data and resources
    ◦ Accessing biomedical information
    ◦ Mapping across biomedical ontologies
  ● Decision support and analytics
    ◦ Value sets and clinical quality measures
  ● Data integration, exchange and semantic interoperability
    ◦ Common data models
    ◦ Fast Healthcare Interoperability Resources (FHIR)

[Bodenreider, YBMI 2008]
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**

- **Not all surface forms represented**
  - Often insufficient for NLP applications
  - Large variation in number of terms per concept across ontologies
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
  • Specific indexing rules
Free cortisol in sepsis and septic shock.

Bendel S¹, Karlsson S, Pettilä V, Loisa P, Varpula M, Ruokonen E; Finnsepsis Study Group.

Abstract

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.

PMID: 18499615 [PubMed - indexed for MEDLINE]
MeSH MEDLINE indexing

MeSH Terms
- Adrenal Cortex Function Tests
- Adrenal Insufficiency/blood*
- Adrenal Insufficiency/drug therapy
- Adrenal Insufficiency/mortality
- Adult
- Biomarkers/blood
- Female
- Finland/epidemiology
- Hospital Mortality
- Humans
- Hydrocortisone/blood*
- Hydrocortisone/therapeutic use
- Kaplan-Meier 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
SNOMED CT/ICD  Coding clinical data

◆ SNOMED CT
  ● Used for clinical documentation
  ● E.g., problem lists

◆ ICD-10-CM
  ● Used for coding clinical data for billing purposes
  ● Other uses of ICD
    - Morbidity and mortality reporting worldwide
  ● Specific coding rules
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 (query expansion)
  - Not always synonymous
- Increase recall without hurting precision

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

Search details
"addison disease"[MeSH Terms] OR
("addison"[All Fields] AND
"disease"[All Fields]) OR "addison
disease"[All Fields] OR
("primary"[All Fields] AND
"hypoadrenalism"[All Fields]) OR
"primary hypoadrenalism"[All Fields]
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 its descendant, Addison disease
Knowledge management

Mapping across biomedical ontologies
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
- SNOMED CT
- OMIM
- MeSH
- Biomedical literature
- NCBI Taxonomy
- GO
- Genome annotations
- FMA
- Anatomy
- Model organisms
- Other subdomains
Integrating subdomains

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

Addison's disease (363732003)

Other subdomains

Clinical repositories

SNOMED CT

OMIM

Genetic knowledge bases

Biomedical literature

Addison Disease (D000224)

Model organisms

NCBI Taxonomy

FMA

GO

Anatomy

Genome annotations
UMLs Source Vocabularies

- 153 families of source vocabularies
  - Not counting translations
- 25 languages
- Broad coverage of biomedicine
  - 9.8M names (normalized)
  - 3.2M concepts
  - ~13M relations among concepts
- Common presentation
Metathesaurus  Basic organization

◆ Concepts
- Synonymous terms are clustered into a concept
- Properties are attached to concepts, e.g.,
  - Unique identifier
  - Definition

◆ Relations
- Concepts are related to other concepts
- Properties are attached to relations, e.g.,
  - Type of relationship
  - Source
Decision support and analytics

Value sets and clinical quality measures
Clinical quality measures (CQMs)

- Measure and track the quality of healthcare services provided by eligible professionals, eligible hospitals and critical access hospitals within our health care system

- Measure many aspects of patient care including:
  - Health outcomes
  - Clinical processes
  - Patient safety
  - Efficient use of healthcare resources
  - Care coordination
  - Patient engagement
  - Population and public health
  - Clinical guidelines
Hemoglobin A1c Test for Pediatric Patients

- Normal glucose levels in blood
  - Low HbA1c concentration

- High glucose levels in blood
  - High HbA1c concentration
Clinical recommendations

1. **American Association of Clinical Endocrinologists (2002):** Recommends that *a glycosylated hemoglobin be performed during an initial assessment and during follow-up assessments*, which should occur at no longer than three-month intervals.

2. **American Diabetes Association (2006):** Recommends *obtaining a glycosylated hemoglobin during an initial assessment and then routinely as part of continuing care*. In the absence of well-controlled studies that suggest a definite testing protocol, expert opinion recommends glycosylated hemoglobin be obtained at least twice a year in patients who are meeting treatment goals and who have stable glycemic control and more frequently (quarterly assessment) in patients whose therapy was changed or who are not meeting glycemic goals.
Hemoglobin A1c Test for Pediatric Patients

# diabetic patients [age 5-17] tested for HbA1c

= 

# diabetic patients [age 5-17]
Hemoglobin A1c Test for Pediatric Patients

Tests for HbA1c

# diabetic patients [age 5-17] tested for HbA1c

# diabetic patients [age 5-17]

- Type 1 or Type 2 diabetes
- Excludes gestational diabetes
- Requires date of birth
Hemoglobin A1c Test for Pediatric Patients

- Type 1 or Type 2 diabetes
- Excludes gestational diabetes

Tests for HbA1c

# diabetic patients [age 5-17] tested for HbA1c

# diabetic patients [age 5-17]

Data element

List of LOINC codes

List of SNOMED CT or ICD 10 codes

- Requires date of birth
## Anatomy of a Clinical Quality Measure

**Population criteria**

- **Initial Patient Population**
  - AND: "Patient Characteristic Birthdate: birth date" \(\geq\) 5 year(s) starts before start of "Measurement Period"
  - AND: "Patient Characteristic Birthdate: birth date" \(\leq\) 17 year(s) starts before start of "Measurement Period"
  - AND: "Diagnosis, Active: Diabetes" starts before or during (MOST RECENT: "Occurrence A of Encounter, Performed: Diabetes Visit" during "Measurement Period")
  - AND: "Encounter, Performed: Diabetes Visit" \(\geq\) 12 month(s) starts before start of "Occurrence A of Encounter, Performed: Diabetes Visit"

- **Denominator**
  - AND: "Initial Patient Population"

- **Denominator Exclusions**
  - AND NOT: "Occurrence A of Diagnosis, Active: Gestational Diabetes" ends before start of "Measurement Period"
  - AND: "Occurrence A of Diagnosis, Active: Gestational Diabetes" starts before or during "Measurement Period"

- **Numerator**
  - AND: 
    
    \[\text{"Laboratory Test, Result: HbA1c Laboratory Test (result)" during "Measurement Period"}\
    
  - **Denominator Exceptions**
  - None

**Data criteria (QDM Data Elements)**

- "Diagnosis, Active: Diabetes" using "Diabetes Grouping Value Set (2.16.840.1.113883.3.464.1003.103.12.1001)"
- "Diagnosis, Active: Gestational Diabetes" using "Gestational Diabetes Grouping Value Set (2.16.840.1.113883.3.464.1003.103.12.1010)"
- "Encounter, Performed: Diabetes Visit" using "Diabetes Visit Grouping Value Set (2.16.840.1.113883.3.464.1003.103.12.1012)"
- "Laboratory Test, Result: HbA1c Laboratory Test" using "HbA1c Laboratory Test Grouping Value Set (2.16.840.1.113883.3.464.1003.198.12.1013)"
- "Patient Characteristic Birthdate: birth date" using "birth date LOINC Value Set (2.16.840.1.113883.3.55.6)"

**Value set = List of LOINC codes for HbA1c tests**
Associated Value Set
Welcome to the NLM Value Set Authority Center (VSAC)

For VSAC announcements, please subscribe to the VSAC Updates listserv.

The Value Set Authority Center (VSAC) is provided by the National Library of Medicine (NLM), in collaboration with the Office of the National Coordinator for Health Information Technology and the Centers for Medicare & Medicaid Services.

The VSAC provides downloadable access to all official versions of vocabulary value sets contained in the 2014 electronic Clinical Quality Measures (eCQMs). Each value set consists of the numerical values (codes) and human-readable names (terms), drawn from standard vocabularies such as SNOMED CT®, RxNorm, LOINC and ICD-10-CM, which are used to define clinical concepts used in clinical quality measures (e.g., patients with diabetes, clinical visit). For information on the eCQMs, visit the eCOI Resource Center.

The content of the VSAC will gradually expand to incorporate value sets for other use cases, as well as for new measures and updates to existing measures.

Viewing or downloading value sets requires a free Unified Medical Language System® Metathesaurus License, due to usage restrictions on some of the codes included in the value sets.

The Data Element Catalog contains the complete list of 2014 CQMs and value set names.
Data integration, exchange and semantic interoperability

Common data models
Clinical data models

- Used in clinical data warehouses
  - Oriented towards analytics
  - Different from the transactional data models of EHR systems
  - Used to normalize data across EHR systems

- Multiple “common” data models
  - OMOP
  - i2b2
  - PCORnet
  - Sentinel
  - CDISC
OMOP – Observational Medical Outcomes Partnership
- Standardized Clinical Data Tables
  - PERSON
  - OBSERVATION_PERIOD
  - SPECIMEN
  - DEATH
  - VISIT_OCCURRENCE
  - PROCEDURE_OCCURRENCE
  - DRUG_EXPOSURE
  - DEVICE_EXPOSURE
  - CONDITION_OCCURRENCE
  - MEASUREMENT
  - NOTE
  - NOTE_NLP (V5.2)
  - OBSERVATION
  - FACT_RELATIONSHIP
- Standardized Health System Data Tables
  - LOCATION
  - CARE_SITE
  - PROVIDER
- Standardized Health Economics Data Tables
  - PAYER_PLAN_PERIOD
  - COST (V5.0.1)
  - VISIT_COST - removed
  - PROCEDURE_COST - removed
  - DRUG_COST - removed
  - DEVICE_COST - removed
- Standardized Derived Elements
  - COHORT
  - COHORT_ATTRIBUTE
  - DRUG_ERA
  - DOSE_ERA
  - CONDITION_ERA
- Standardized Vocabularies
  - CONCEPT
  - VOCABULARY
  - DOMAIN
  - CONCEPT_CLASS
  - CONCEPT_RELATIONSHIP
  - RELATIONSHIP
  - CONCEPT_SYNONYM
  - CONCEPT_ANCESTOR
  - SOURCE_TO_CONCEPT_MAP
  - DRUG_STRENGTH
  - COHORT_DEFINITION
  - ATTRIBUTE_DEFINITION
- Standardized meta-data
  - CDM_SOURCE
i2b2

- i2b2 – Informatics for Integrating Biology & the Bedside
- Originally developed by the i2b2 National Center for Biomedical Computing (2004-2013)
  - Now i2b2 tranSMART Foundation
- Platform to support translational research
- Widely adopted worldwide

https://www.i2b2.org/
i2b2 data model – original “star schema”
i2b2-OMOP convergence

◆ i2b2 on OMOP
  • Supports query formulation against an OMOP-compliant data source through i2b2 tools
PCORnet – National Patient-Centered Clinical Research Network

Initiative of the Patient-Centered Outcomes Research Institute (PCORI)
- Funded through the Patient Protection and Affordable Care Act of 2010

“designed to make it faster, easier, and less costly to conduct clinical research”

Made up of
- 13 Clinical Data Research Networks (CDRNs)
- 20 Patient-Powered Research Networks (PPRNs)
### PCORnet Common Data Model v3.0

#### Fundamental basis
- **DEMOGRAPHIC**
  - PATID
  - BIRTH_DATE
  - BIRTH_TIME
  - SEX
  - HISPANIC
  - RACE
  - BIOBANK_FLAG

Data captured from processes associated with healthcare delivery, registry activity, or directly from patients.

#### Associations with PCORnet clinical trials
- **PCORNET_TRIAL**
  - PATID
  - TRIALID
  - PARTICIPANTID
  - TRIAL_SITEID
  - TRIAL_ENROLL_DATE
  - TRIAL_END_DATE
  - TRIAL_WITHDRAW_DATE
  - TRIAL_INVITE_CODE

#### Process-related data
- **HARVEST**
  - NETWORKID
  - NETWORK_NAME
  - DATAMART_NAME
  - DATAMARTPLATFORM
  - CDM_VERSION
  - DATAMART_CLAIMS
  - DATAMART_EHR
  - BIRTH_DATE_MGMT
  - ENR_START_DATE_MGMT
  - ENR_END_DATE_MGMT
  - ADMIT_DATE_MGMT
  - DISCHARGE_DATE_MGMT
  - SPECIMEN_DATE_MGMT
  - RESULT_DATE_MGMT
  - MEASURE_DATE_MGMT
  - REPORT_DATE_MGMT
  - RESOLVE_DATE_MGMT
  - REFRESH_DEMOGRAPHIC_DATE
  - REFRESH_ENROLLMENT_DATE
  - REFRESH_ENCOUNTER_DATE
  - REFRESH_DIAGNOSIS_DATE
  - REFRESH_PROCEDURE_DATE
  - REFRESH_VITAL_DATE
  - REFRESH_DISPENSING_DATE

#### New to v3.0
- **ENCOUNTER**
  - PATID
  - ENCLUSTERID (optional)
  - ENC_TYPE

#### Data captured within multiple contexts: healthcare delivery, registry activity, or directly from patients
- **PRO_CM**
  - PATID
  - PRO_CM_ID
  - PRO_ITEM
  - PRO_LOINC
  - PRO_DATE
  - PRO_TIME
  - PRO_RESPONSE
  - PRO_METHOD
  - PRO_MODE
  - PRO_CAT

- **PROCEDURES**
  - PATID
  - ENCOUNTERID
  - PRO_PROVIDED
  - PRO_PROVIDER
  - PRO_TIME
  - PRO_DATE
  - PRO_LOCATION
  - PRO_TYPE
  - PRO_SERVICE

- **DIAGNOSIS**
  - PATID
  - ENCOUNTERID
  - ENC_TYPE (replicated)
  - ADMIT_DATE (replicated)
  - PROVIDER
  - ITEM
  - UNIT
  - MODIFIER

- **LAB_RESULT_CM**
  - PATID
  - ENCOUNTERID
  - LAB_NAME
  - SPECIMEN_SOURCE
  - LOINC
  - RESULT
  - RESULT_UNIT
  - RESULT_DATE

### References
- [PCORnet Common Data Model](http://www.pcornet.org/pcornet-common-data-model/)

_NLM_
Sentinel

- Initiative of the Food and Drug Administration (FDA)
- Effort to create a national electronic system for monitoring the performance of FDA-regulated medical products (drugs, vaccines, and other biologics)
- Develop a system to obtain information from existing electronic health care data from multiple sources to assess the safety of approved medical products
- Distributed dataset reached 100 lives in 2011
## Sentinel Common Data Model

### List of Tables

<table>
<thead>
<tr>
<th>Table Name</th>
<th>Source</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Enrollment</td>
<td>Created by Data Partners using Data Partner data.</td>
<td>The SCDM Enrollment Table has a start/stop structure that contains one record per continuous enrollment period. Members with medical coverage, drug coverage, or both should be included. A unique combination of PatiID, Enr_Start, Enr_End, MedCov, DrugCov, and Chart identifies a unique record. A break in enrollment (of at least one day) or a change in either the medical or drug coverage variables should generate a new record.</td>
</tr>
<tr>
<td>2. Demographic</td>
<td></td>
<td>The SCDM Demographic Table contains one record per PatiID with the most recent information on Birth_Date, Sex, Race/Ethnicity, and Zip Code.</td>
</tr>
<tr>
<td>3. Dispensing</td>
<td>Created by Data Partners using Data Partner data.</td>
<td>The SCDM Outpatient Pharmacy Dispensing Table contains one record per unique combination of PatiID, NDC, and RxDate. Each record represents an outpatient pharmacy dispensing. Rollback transactions and other adjustments should be processed before populating this table.</td>
</tr>
<tr>
<td>4.1 Encounter</td>
<td>Created by Data Partners using Data Partner data.</td>
<td>The SCDM Encounter Table contains one record per PatiID and EncounterID. Each encounter should have a single record in the SCDM Encounter Table. Each diagnosis and procedure recorded during the encounter should have a separate record in the Diagnosis or Procedure Tables. Multiple visits to the same provider on the same day should be considered one encounter and should include all diagnoses and procedures that were recorded during those visits. Visits to different providers on the same day, such as a physician appointment that leads to a hospitalization, should be considered multiple encounters. Rollback transactions and other adjustments should be processed before populating this table.</td>
</tr>
<tr>
<td>4.2 Diagnosis</td>
<td>Created by Data Partners using Data Partner data.</td>
<td>The SCDM Diagnosis Table contains one record per unique combination of PatiID, EncounterID, DX, and DX_CodeType. This table should capture all uniquely recorded diagnoses for all encounters.</td>
</tr>
<tr>
<td>4.3 Procedure</td>
<td>Created by Data Partners using Data Partner data.</td>
<td>The SCDM Procedure Table contains one record per unique combination of PatiID, EncounterID, and PX_CodeType. This table should capture all uniquely recorded procedures for all encounters.</td>
</tr>
<tr>
<td>5.1 Death</td>
<td>Created by Data Partners using Data Partner data.</td>
<td>The SCDM Death Table contains one record per PatiID. When legacy data have conflicting reports, make a local determination as to which to use. There is typically a 1-2 year lag in death registry data.</td>
</tr>
<tr>
<td>5.2 Cause of Death</td>
<td>Created by Data Partners using Data Partner data.</td>
<td>The SCDM Cause of Death Table contains one record per unique combination of PatiID and COD. Legacy data have conflicting reports, please make a local determination as to which to use. There is typically a 1-2 year lag in death registry data.</td>
</tr>
<tr>
<td>6.1 Laboratory Result</td>
<td>Created by Data Partners using Data Partner data.</td>
<td>The SCDM Laboratory Result Table contains one record per result/entry. Only include results if Data Partners are strongly encouraged to review the comprehensive Sentinel Common Data Model Laboratory Result Table Documentation for details on how to populate each variable.</td>
</tr>
</tbody>
</table>
CDISC

- CDISC – Clinical Data Interchange Standards Consortium
- “develop and support global, platform-independent data standards that enable information system interoperability to improve medical research and related areas of healthcare”
- Set of standards required by FDA for regulatory submissions (clinical research)

https://www.cdisc.org/
Common data models in action

Characterizing treatment pathways at scale using the OHDSI network

George Hripcsak\textsuperscript{a,b,c,1}, Patrick B. Ryan\textsuperscript{c,d}, Jon D. Duke\textsuperscript{c,e}, Nigam H. Shah\textsuperscript{c,f}, Rae Woong Park\textsuperscript{c,g}, Vojtech Huser\textsuperscript{c,h}, Marc A. Suchard\textsuperscript{c,i,j,k}, Martijn J. Schuemie\textsuperscript{c,d}, Frank J. DeFalco\textsuperscript{c,d}, Adler Perotte\textsuperscript{a,c}, Juan M. Banda\textsuperscript{c,f}, Christian G. Reich\textsuperscript{c,l}, Lisa M. Schilling\textsuperscript{c,m}, Michael E. Matheny\textsuperscript{c,n,o}, Daniella Meeker\textsuperscript{c,p,q}, Nicole Pratt\textsuperscript{c,r}, and David Madigan\textsuperscript{c,s}

www.pnas.org/cgi/doi/10.1073/pnas.1510502113

PNAS | July 5, 2016 | vol. 113 | no. 27 | 7329–7336
A Diabetes

- Metformin
- pioglitazone
- sitagliptin
- Glipizide
- glimepiride
- Gliclazide
- Glyburide
- rosiglitazone
- Insulin, Glargine, Human
- exenatide
- Insulin, Aspart, Human
- liraglutide
- saxagliptin
- Insulin, Lispro, Human
- Glucose
- Insulin, Isophane, Human
Data integration, exchange and semantic interoperability

Fast Healthcare Interoperability Resources (FHIR)
Fast Healthcare Interoperability Resources

◆ New standard for exchanging healthcare information electronically
◆ Developed by HL7 FHIR foundation
◆ Based on resources
  ● Basic building blocks of information (patient, condition, procedure, practitioner)
  ● Can be extended as needed
◆ Supports 4 different paradigms for exchange: the RESTful API, Messaging, Documents, and Services

http://www.fhir.org/
Clinical Terminology

Summary
Summary

◆ How clinical features are reflected in disease names

◆ Structure of the main clinical ontologies used
  ● ICD, SNOMED CT, LOINC, RxNorm

◆ Use cases for biomedical ontologies
  ● Knowledge management
  ● Decision support and analytics
  ● Data integration, exchange and semantic interoperability
Topics not discussed

- Semantic Web, URIs, Linked Data
- Ontology creation, Protege
- Accessing terminology resources (APIs)
- Ontology repositories
  - [UMLS], NCBO BioPortal, EBI Ontology Lookup Service
- NLP, named entity recognition, MetaMap
- Mapping local terms to standard terminologies
- OBO ontologies, OBO Foundry
- Coordinated development of ontologies, harmonization
- Boundary between terminology and information model
- […]
Medical Ontology Research

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

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

U.S. National Library of Medicine