Comparability and Consistency of Clinical Data: The ICD11 initiative

Christopher G. Chute, MD DrPH
Bloomberg Distinguished Professor of Health Informatics
Professor of Medicine, Public Health, and Nursing
Chief Research Information Officer
Johns Hopkins Medicine, Baltimore, MD, USA

Chair, International Classification of Disease Revision, WHO
Former Chair, ISO TC215 Health Informatics

ISPOR
Singapore, 5 Sept 2016
Origins of Big Science

Astronomy

Sloan Digital Sky Survey III – DR9

- Images
- Spectra
- Object catalog
- Metadata

Total area of imaging: 31,637 square degrees
Image field size: 1361x2048 pixels
Number fields: 938,046 (excluding supernovae runs)
Catalog objects: 1,231,051,050
Number of unique, primary sources:
- Stars: 260,562,744
- Galaxies: 208,478,448
- Unknown: 12,682

Number of unique, primary sources:
- Total: 469,053,874
- Stars: 260,562,744
- Galaxies: 208,478,448
- Unknown: 12,682
Rutherford “Table-top” Experiment

That Higgs Boson

- 600 institutions
- 10,000 scientists
  - 2 Nobel prizes
- 800 trillion collisions
- 200PB of data = $2 \times 10^{17}$ bytes of data

Boarding on an astronomical number in its own right!
- $13.25 \times 10^9$ USD
Dimensionality of Higgs “Big Data”

- Mass/Energy
- Direction
- Charge

- Medicine is more complicated than that

---

Dimensionality of Big Data

- Broad
  - Small amounts of data; Huge number observations
  - National Claims data
- Deep
  - Large amounts of data; Few observation
  - NGS Complete Genome
- Rich
  - Broad and Deep
  - Clinical Phenotyping data (EMRs)
    - Labs, Vitals, Exam, Waveform, Images, Omics, …
    - Social, environmental, diet,
Will Big Data Save Us?

Genet Med 15: 802-809; Oct, 2013

Some experiences and opportunities for big data in translational research

Christopher G. Chute, MD, DrPH\(^1\), Mollie Ullman-Cullere, MS, MSE\(^2\), Grant M. Wood, BS\(^3\), Simon M. Lin, MD\(^4\), Min He, PhD\(^5\) and Jyotishman Pathak, PhD\(^1\)

Health care has become increasingly information intensive. The advent of genomic data, integrated into patient care, significantly accelerates the complexity and amount of clinical data. Translational research in the present day increasingly embraces new biotechnological discovery in this data-intensive world, thus entering the domain of “big data.” The Electronic Medical Records and Genomics consortium has taught us many lessons, while simultaneously advances in commodity computing methods enable the academic community to affordably manage and process big data. Although great promise can emerge from the adoption of big data methods and philosophy, the heterogeneity and complexity of clinical data, in particular, pose additional challenges for big data inference and clinical application. However, the ultimate comparability and consistency of heterogeneous clinical information sources can be enhanced by existing and emerging data standards, which promise to bring order to clinical data chaos. Meaningful Use data standards in particular have already simplified the task of identifying clinical phenotyping patterns in electronic health records.

Key Words: clinical data representation; big data; genomics; health information technology standards

Comparable and Consistent Semantics as Crucial Requirement

Without Terminology/Ontology Standards...

• Health Data is non-comparable
• Health Systems cannot Interchange “Data”
• Secondary Uses (Research) are not practical
• Big Data methodologies cannot leverage epidemiologic principles for observational data
  • Adjustment for confounding
    • Stratification
    • Multivariate models
    • Machine learning features
From Practice-based Evidence to Evidence-based Practice

Data
- Clinical Databases
- Registries et al.

Inference
Comparability and Consistency

Standards
- Terminologies & Data Models

Medical Knowledge Management

Decision Support
- Expert Systems
- Clinical Guidelines

Patient Encounters

Foundations for Learning Health System

Natural and Political Observations
Mentioned in a following Index and made upon the Bills of Mortality.

By Capt. John Grant, Fellow of the Royal Society.

With reference to the Government, Religion, Trade, Growth, Air, Disease, and the several Changes of the said City.


London, Printed by John Martyn, and James Allestry, Printers to the Royal Society, and are to be sold at the sign of the Red in St. Pauls Church-yard. MDCLXV.
"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."

International Classification of Disease
ICD11 Use Cases

• Scientific consensus of *clinical phenotype*
• Public Health Surveillance
  • Mortality
  • Public Health Morbidity
• Clinical data aggregation
  • Metrics of clinical activity
  • Quality management
    • Patient Safety
  • Financial administration
    • Case mix
    • Resource allocation

Traditional Hierarchical System
ICD-10 and family
The ICD11 Foundation Component

a Semantic Network

ICD Concept Title
Unique Identifier (URI)
Fully Specified Name
Preferred Name
Synonyms
Classification
Properties
Parents
Type
Use and Linearization(s)
Textual Definition(s)
Terms
Base Index Terms
Inclusion Terms
Exclusions

Body System/Structure
Manifestation
Causal
Etiology
Genomics
Agents
Temporal Severity
Functioning Properties
Specific Condition
Gender
Life Cycle
Treatment
Diagnostic Criteria

Algorithmic Serialization of the Foundation Component into a Linearization

Mutually Exclusive And Exhaustive
Linearizations for multiple use-cases
Morbidity, Mortality, Quality, ...

Relationship with IHTSDO
SNOMED content

- IHT (SNOMED) will require high-level nodes that aggregate more granular data
  - Use-cases include mutually exclusive, exhaustive,…
  - Sounds a lot like ICD
- ICD-11 will require lower level terminology for value sets which populate content model
  - Detailed terminological underpinning
  - Sounds a lot like SNOMED
- Memorandum of Agreement – July 2010!
  - WHO right to use for authoring and interpretation
Potential Future States (2007)

ICD-11

SNOMED

Ghost SNOMED

Ghost ICD

Joint ICD-IHTSDO Effort

ICD-11

SNO
**Common Ontology**

- Joint effort between WHO and IHTSDO
- A subset of SNOMED CT
- Provides semantic anchoring of ICD11 Foundation Component
  - Semantic backbone
- All ICD11 Foundation Component elements will be defined by “query expressions” against the Common Ontology

---

**ICD 11 Linkage Queries**

Links between Foundation Component and Linearizations

All linearization entities are represented as queries against Common Ontology

**Morbidity Linearization**

**Residual Categories**

**NEC**  **NOS**

**Morbidity**

“Hypertension excluding Pregnancy”

```sparql
SELECT ?CN WHERE
(?CN SubclassOf Hypertension)
MINUS
(?CN SubclassOf Disorders of Pregnancy)
```
Where is This Going?

• Biomedical practice and research are data, information, and knowledge intensive

• Comparable and consistent data representation are pre-requisite for efficient clinical analytics

• Data standards are needed to support comparable and consistent information

• ICD-11 promises to become a multi-use standard for the 21st century