Processing Healthcare Data

Guest Lecture

Olivier Bodenreider
Acting Director, Lister Hill National Center for Biomedical Communications
National Institutes of Health, Bethesda, Maryland, USA
Disclaimer

The views and opinions expressed do not necessarily state or reflect those of the U.S. Government, and they may not be used for advertising or product endorsement purposes.
Outline

• Introduction to the Lister Hill National Center for Biomedical Communications
• Interoperability and biomedical terminologies
• Analyzing Opioid Prescriptions in Medicare
Introduction to the Lister Hill National Center for Biomedical Communications
National Library of Medicine

• Largest biomedical library in the world
• Started in 1836 as a small collection of medical books and journals in the office of the United States Army Surgeon General
• Part of the National Institutes of Health since 1962
• Flagship products and services (among many others)
  • PubMed/MEDLINE
  • ClinicalTrials.gov
  • Unified Medical Language System (UMLS)
National Library of Medicine

• Curating data, not just (dusty) books
  • Biomedical literature
  • Gene sequences
  • Clinical trials
  • Information for lay public
  • [...]

• Research, not just products and services
  • NLM Intramural Research program
    • Computational Biology – National Center for Biotechnology Information (NCBI)
    • Computational Health – Lister Hill National Center for Biomedical Communications
  • NLM Intramural Training program

• Extramural programs (grants)
Lister Hill National Center for Biomedical Communications (LHC)

• With NCBI, one of the two research & development centers of NLM

• Established in 1968

• Initially focused on biomedical communications
  • Communication networks applied to health
  • Audiovisual technologies in health applications
  • Use of new technologies for health education

• Current reorganization around health informatics
  • Clinical data science: Interoperable data, scalable methods and translation of discovery into operations
LHC within NLM and NIH (FY2020)

- **NIH**
  - $41.46 billion
  - 20,000 employees

- **NLM**
  - $456 million
  - 1700 employees

- **LHC**
  - $19 million (estimated)
  - 100 employees
4 research areas
LHC research and development activities

• Natural Language Processing
  • Identifying biomedical concepts and relations in clinical text / literature
  • Clinical question answering

• Image processing
  • Application of machine learning/deep learning techniques to imaging datasets to support diagnostics

• Health information standards
  • Terminology standards (UMLS, SNOMED CT, MeSH, RxNorm, LOINC, …)
  • Information model standards (common data models, FHIR – Fast Healthcare Interoperability Resource)

• Health data-powered discovery
  • Getting insights from large observational databases (EHR and claims data)
LHC Postdoctoral Fellows

- N=86 (over the past 10 years)
- Demographics
  - 40% women, 60% men
- Career after leaving NLM
  - Academia (50%)
  - Industry (34%)
  - Other (16%)
Interoperability among biomedical terminologies
Many biomedical terminologies

• Different purposes
  • Clinical documentation – fine grained
  • Morbidity and mortality statistics – classification (avoid double-counting)
  • Indexing/retrieval – abstraction
  • Text mining – lexical variation

• Developed independently
  • Standard Development Organizations
  • No standard for developing standards
  • Different funding mechanisms
  • Different legacy products
Internal Classification of Diseases
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]

[...]

15
Practical interoperability issues

• Data integration
  • Analyze datasets coded with different terminologies
    • Biomedical literature indexed with MeSH
      • Pancreatic neoplasm (D010190)
    • Healthcare utilization data (e.g., HCUP – Healthcare Cost and Utilization Project) coded with Clinical Classifications Software (CCS)
      • Cancer of pancreas (17)
  • Using a specific terminology to aggregate data
    • Drugs coded with [A]NDA ([Abbreviated] New Drug Application) or NDC (National Drug Code)
    • Analysis based on the ATC (Anatomical-Therapeutic-Chemical) drug classification
Degrees of semantic interoperability

- **Synonymy**
  - Equivalence between terms (or concepts)
  - Myocardial infarction ↔ Heart attack

- **Mapping**
  - Closest term for the source term in the target terminology
  - Lipitor → Atorvastatin

- **Closest ancestor**
  - Closest term in the target terminology among the ancestors in the source
  - Pancreatic cancer → Pancreatic neoplasm

- **Post-coordination**
  - One term equivalent to the combination of several terms in the target terminology
  - Diabetic nephropathy → Nephropathy + Diabetes mellitus
Integrating datasets

- Clinical repositories
- Other subdomains
- SNOMED CT
- OMIM
- MeSH
- Biomedical literature
- Genetic knowledge bases
- GO
- FMA
- Genome annotations
- Anatomy
- Model organisms
- NCBI Taxonomy

...
Point-to-point mappings are impractical

- Clinical repositories
- Genetic knowledge bases
- Biomedical literature
- Genome annotations
- Other subdomains
- Model organisms
- Anatomy
Integration through a reference (e.g., UMLS)

Addison's disease (363732003)

Other subdomains

Model organisms

NCBI Taxonomy

FMA

Anatomy

SNOMED CT

UMLS C0001403

Clinical repositories

OMIM

Genetic knowledge bases

MeSH

Biomedical literature

Addison Disease (D000224)

Genome annotations

Clinical repositories

Genome annotations

Other subdomains

Model organisms

NCBI Taxonomy

FMA

Anatomy

SNOMED CT

UMLS C0001403

Clinical repositories

OMIM

Genetic knowledge bases

MeSH

Biomedical literature

Addison Disease (D000224)

Genome annotations
Semantic interoperability through UMLS

• Synonymy
  • *Synonymous terms clustered into the same ULMS concept*
  • Myocardial infarction ↔ Heart attack

• Mapping
  • *Existing mapping tables integrated into UMLS (e.g., ICD10 to SNOMED CT)*
  • Lipitor → Atorvastatin

• Closest ancestor
  • *Hierarchical relations are recorded in UMLS and can be navigated*
  • Pancreatic cancer → Pancreatic neoplasm

• Post-coordination
  • *Logical definitions for concepts re recorded in UMLS (whenever available)*
  • Diabetic nephropathy → Nephropathy + Diabetes mellitus
Pancreatic cancer → Pancreatic neoplasm

Legend

- UMLS
- MeSH
- SNOMED CT
- CCS

C0346647

Malignant tumor of pancreas [363418001]

Pancreatic cancer [17]

C0030297

Pancreatic Neoplasms [D010190]

Neoplasm of pancreas [126859007]

MeSH

Neoplasm of pancreas [126859007]

SNOMED CT

Malignant tumor of pancreas [363418001]

Pancreatic cancer [17]
Analyzing Opioid Prescriptions in Medicare
Medicare Part D dataset

• Main variables in the Drug Event File
  • Beneficiary information (ID and demographics)
  • Date on which the prescription was filled
  • Drug: identified by NDC (11-digit format)
  • Quantity Dispensed
  • Days Supply
  • Cost information

• Related information for the NDC (provided by First Databank)
  • Brand name, generic name, strength, dosage form code, and dosage form description

Use case: Analysis of opioid prescriptions

• Identify prescriptions corresponding to opioids in the Medicare part D dataset
• For each opioid drug, calculate the trend of dispensation over time (“number of prescriptions”)
• For all opioids, calculate the trend of total (or daily) dose dispensed in oral morphine milligram equivalents
Analysis of opioid prescriptions – How to?

• Identify prescriptions corresponding to opioids in the Medicare part D dataset
  • Use ATC classes to identify opioid drugs
  • Link NDC codes (Medicare) to ATC codes through RxNorm
• For each opioid drug, calculate the trend of dispensation over time ("number of prescriptions")
• For all opioids, calculate the trend of total (or daily) dose dispensed in oral morphine milligram equivalents
Identifying opioid drugs from a drug class

• ATC – Anatomical Therapeutic Chemical drug classification system

<table>
<thead>
<tr>
<th>ATC code</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>N02AA01</td>
<td>morphine</td>
</tr>
<tr>
<td>N02AA02</td>
<td>opium</td>
</tr>
<tr>
<td>N02AA03</td>
<td>hydromorphone</td>
</tr>
<tr>
<td>N02AA04</td>
<td>nicomorphine</td>
</tr>
<tr>
<td>N02AA05</td>
<td>oxycodone</td>
</tr>
<tr>
<td>N02AA06</td>
<td>dihydromorphone</td>
</tr>
<tr>
<td>N02AA07</td>
<td>papaveretum</td>
</tr>
<tr>
<td>N02AA51</td>
<td>morphine, combinations</td>
</tr>
<tr>
<td>N02AA53</td>
<td>hydromorphone and naloxone</td>
</tr>
<tr>
<td>N02AA55</td>
<td>oxycodone and naloxone</td>
</tr>
<tr>
<td>N02AA56</td>
<td>oxycodone and nalbuphine</td>
</tr>
<tr>
<td>N02AA58</td>
<td>dihydromorphone, combinations</td>
</tr>
<tr>
<td>N02AA59</td>
<td>codeine, combinations excl. psycholeptics</td>
</tr>
<tr>
<td>N02AA79</td>
<td>codeine, combinations with psycholeptics</td>
</tr>
</tbody>
</table>
Linkages among drug entities

oxycodone [N02AA05]

oxyCODONE [7804]

Abuse-Deterrent 12 HR oxyCODONE Hydrochloride 10 MG Extended Release Oral Tablet [1860157]

12 HR OxyCONTIN 10 MG Extended Release Oral Tablet [1049504]

Nervous System
N02 Analgesics
N02A Opioids
N02AA Natural opium alkaloids

ATC code  Name          DDD   U      Adm.R.  Note
N02AA05  oxycodone    75 mg  30 mg  P
Why converting to MME?

• Opioids have widely different potency levels
  • Fentanyl is about 100 times more potent than morphine

• Difficult to
  • Compare doses across drugs
  • Compare doses over time for multiple drugs
  • Aggregate results

• Reference: 1 mg of morphine administered orally

• Use case: How do these two drugs compare?
  • 12 HR OxyCONTIN 10 MG Extended Release Oral Tablet (twice a day)
  • 72 HR fentaNYL 0.012 MG/HR Transdermal System
MME conversion factor

- Conversion factor for each drug
  - Available from CMS
  - Compiled from CDC data

Example #1

• Drug: NDC = 59011041010
  • 12 HR OxyCONTIN 10 MG Extended Release Oral Tablet [1049504]
  • Ingredient: Oxycodone
  • Strength: 10 MG

• Dispensation information
  • Quantity Dispensed: 20
  • Days Supply: 10

• MME conversion factor: 1.5

\[ MME(mg) = 10 \times \frac{20}{10} \times 1.5 = 30 \, mg \]

• MME/day: