Columbia Business School
AT THE VERY CENTER OF BUSINESS™

February 11, 2020
Economics of Healthcare and Pharmaceuticals
SP2020

Processing Healthcare Data

Guest Lecture

Olivier Bodenreider

Acting Director, Lister Hill National Center for Biomedical Communications

National Institutes of Health, Bethesda, Maryland, USA

NIH U.S. National Library of Medicine

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

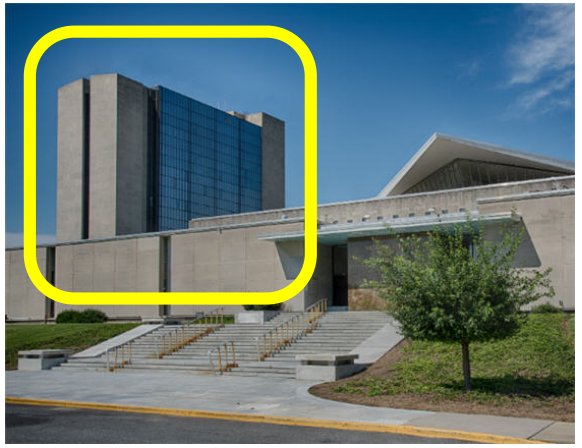
National Library of Medicine

- 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)

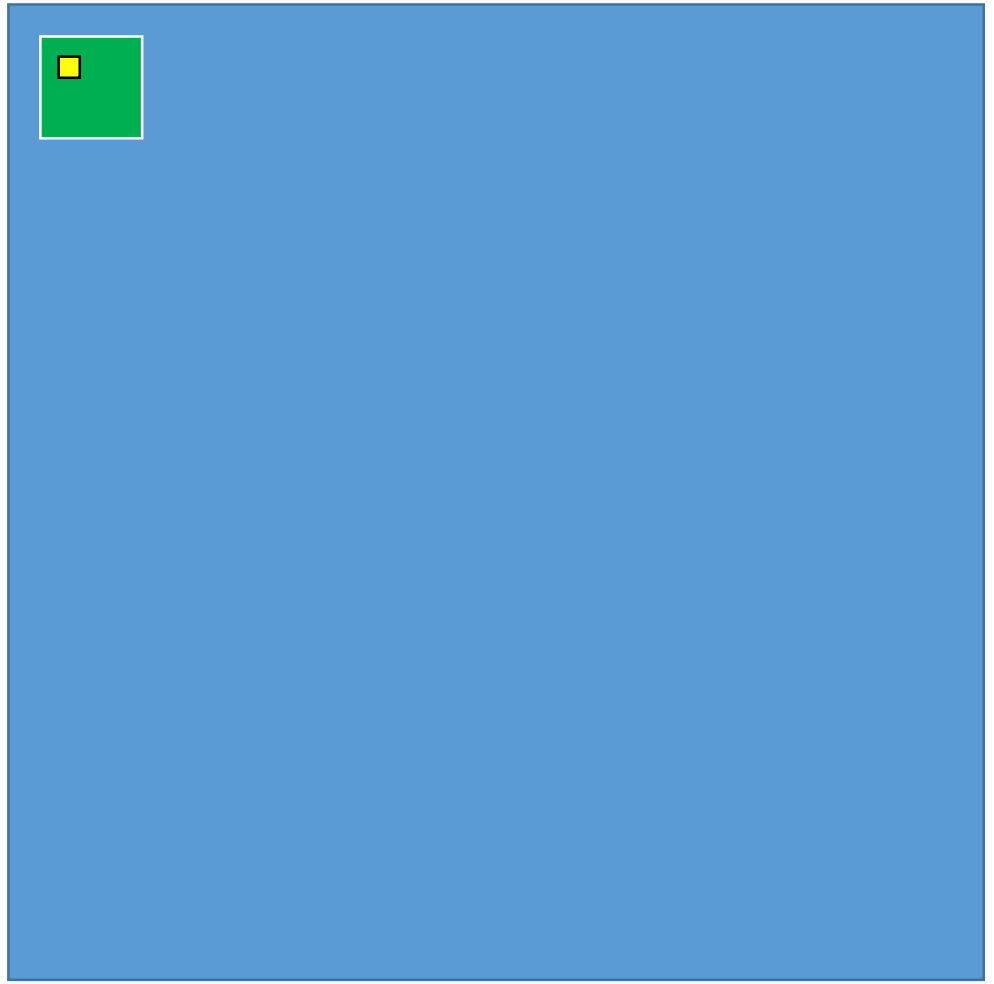


- With NCBI, one of the two research & development centers of NLM
- Established in 1968
- Initially focused on biomedical communications
  - Communication networks applied to health
  - Audiovisual technologies in health applications
  - Use of new technologies for health education

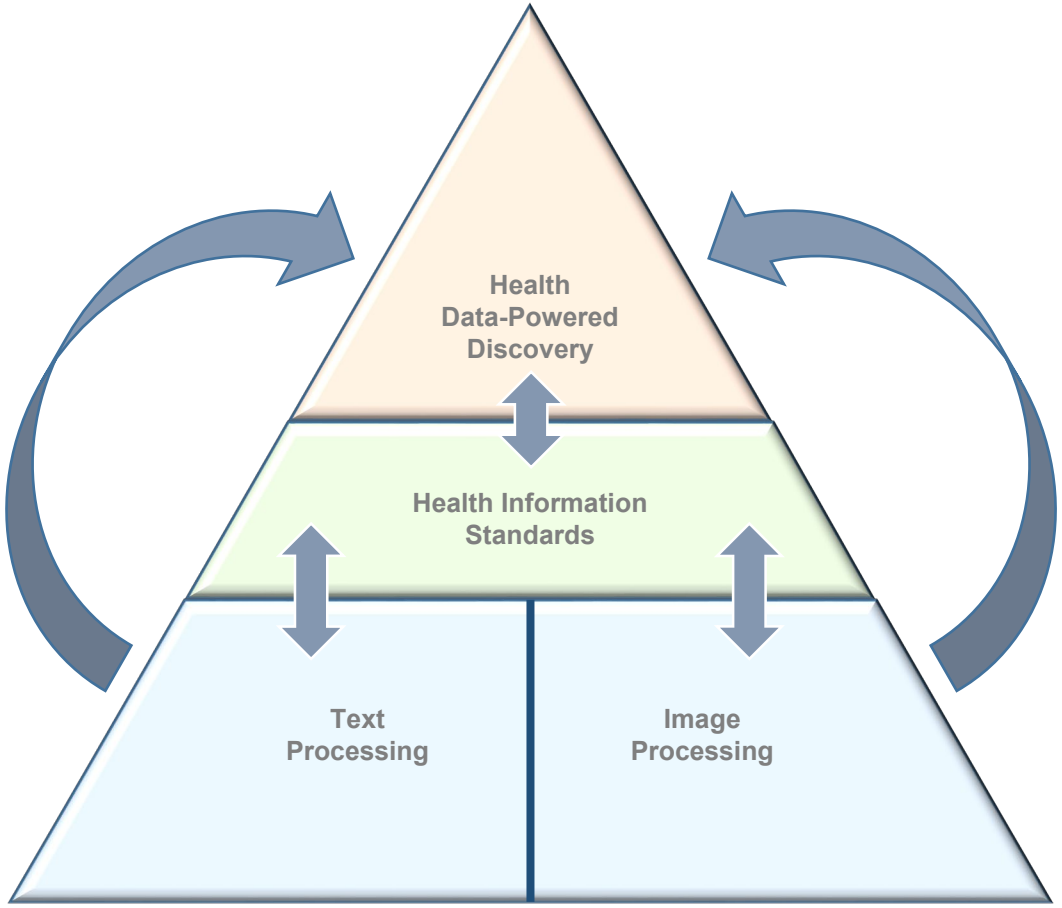
- Current reorganization around health informatics
  - Clinical data science: Interoperable data, scalable methods and translation of discovery into operations

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

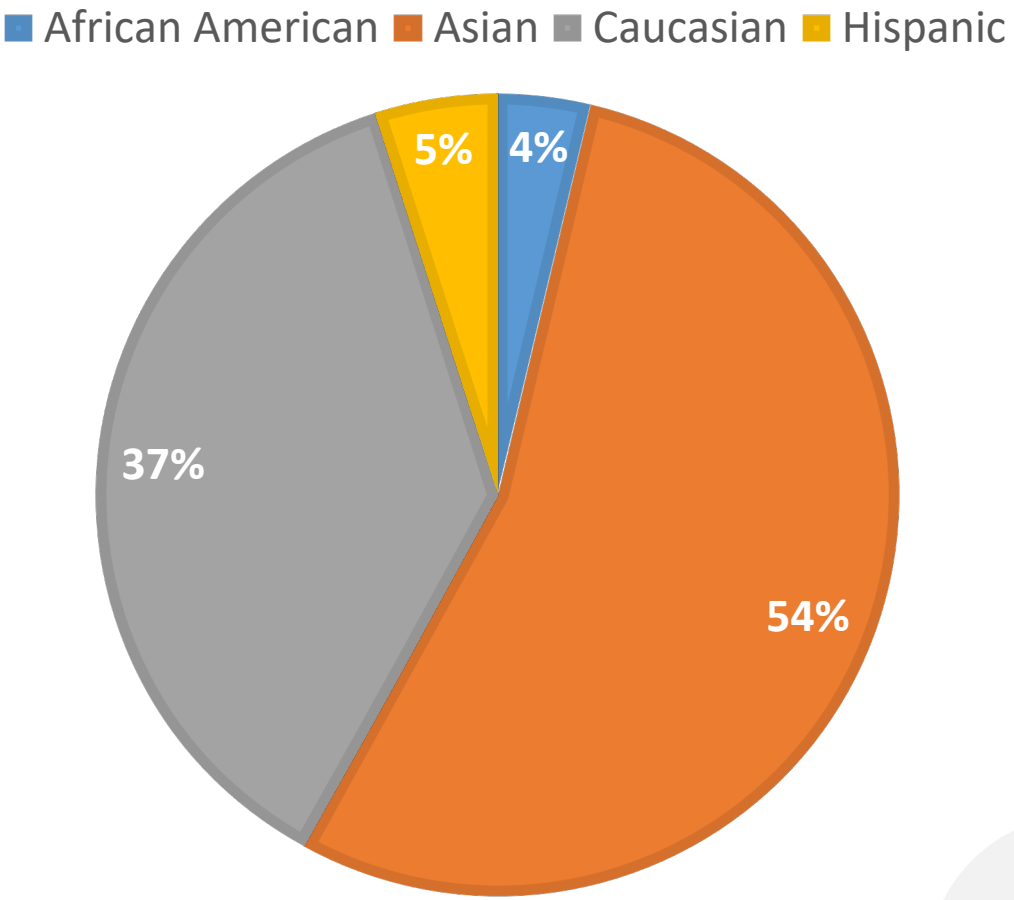


LHC research and development activities

- 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%)

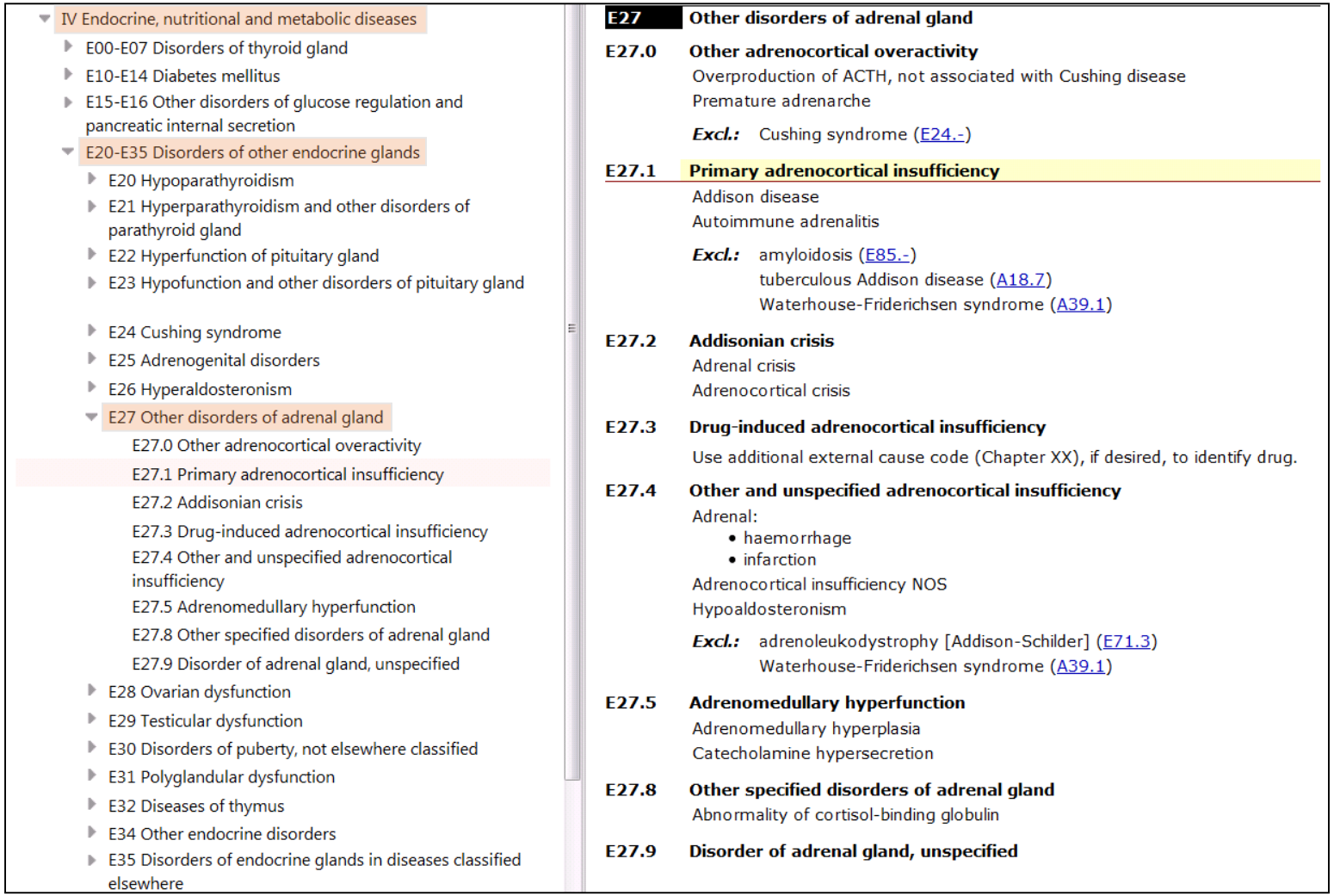


Interoperability among biomedical terminologies

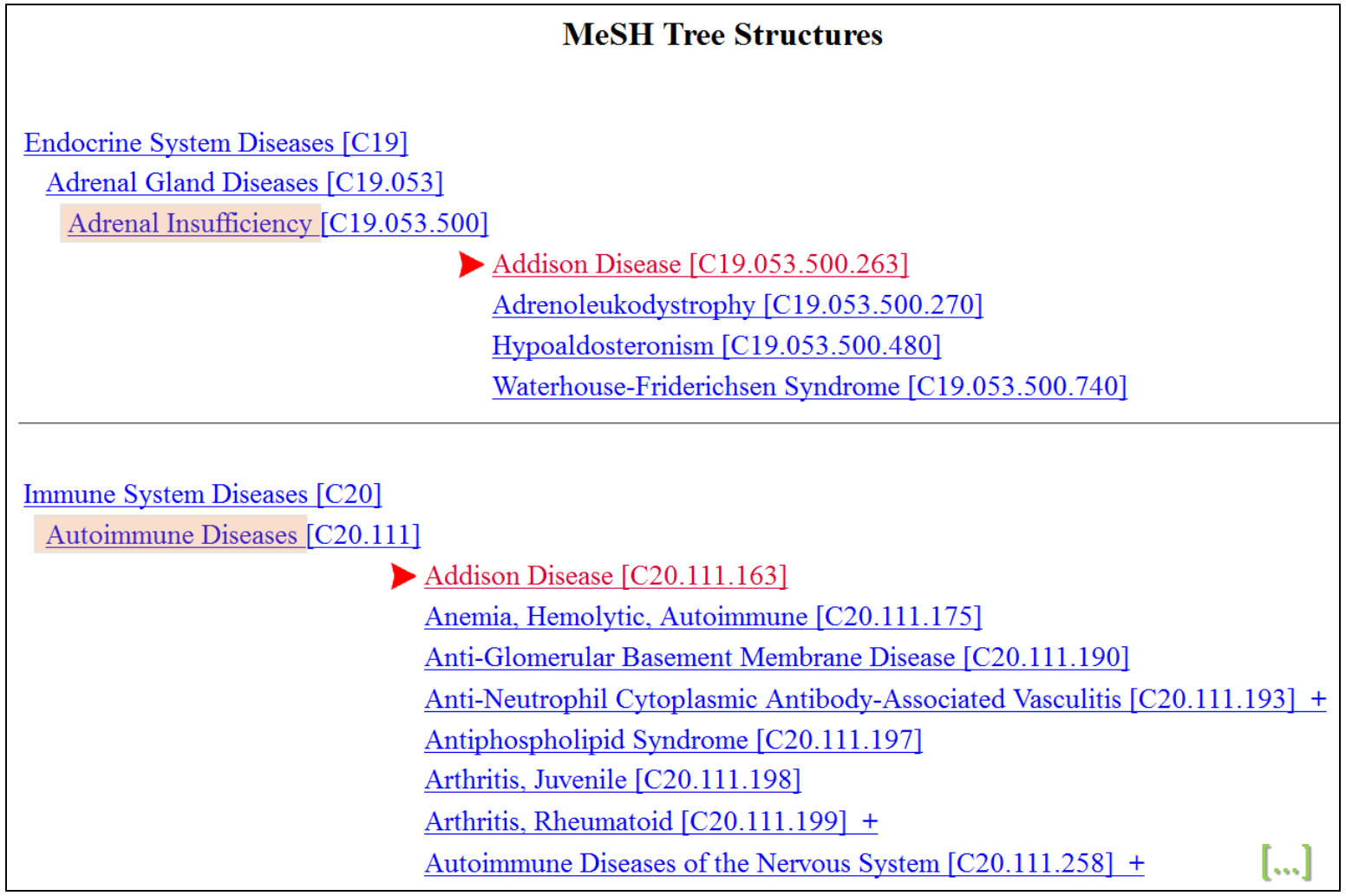
Many biomedical terminologies

- 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



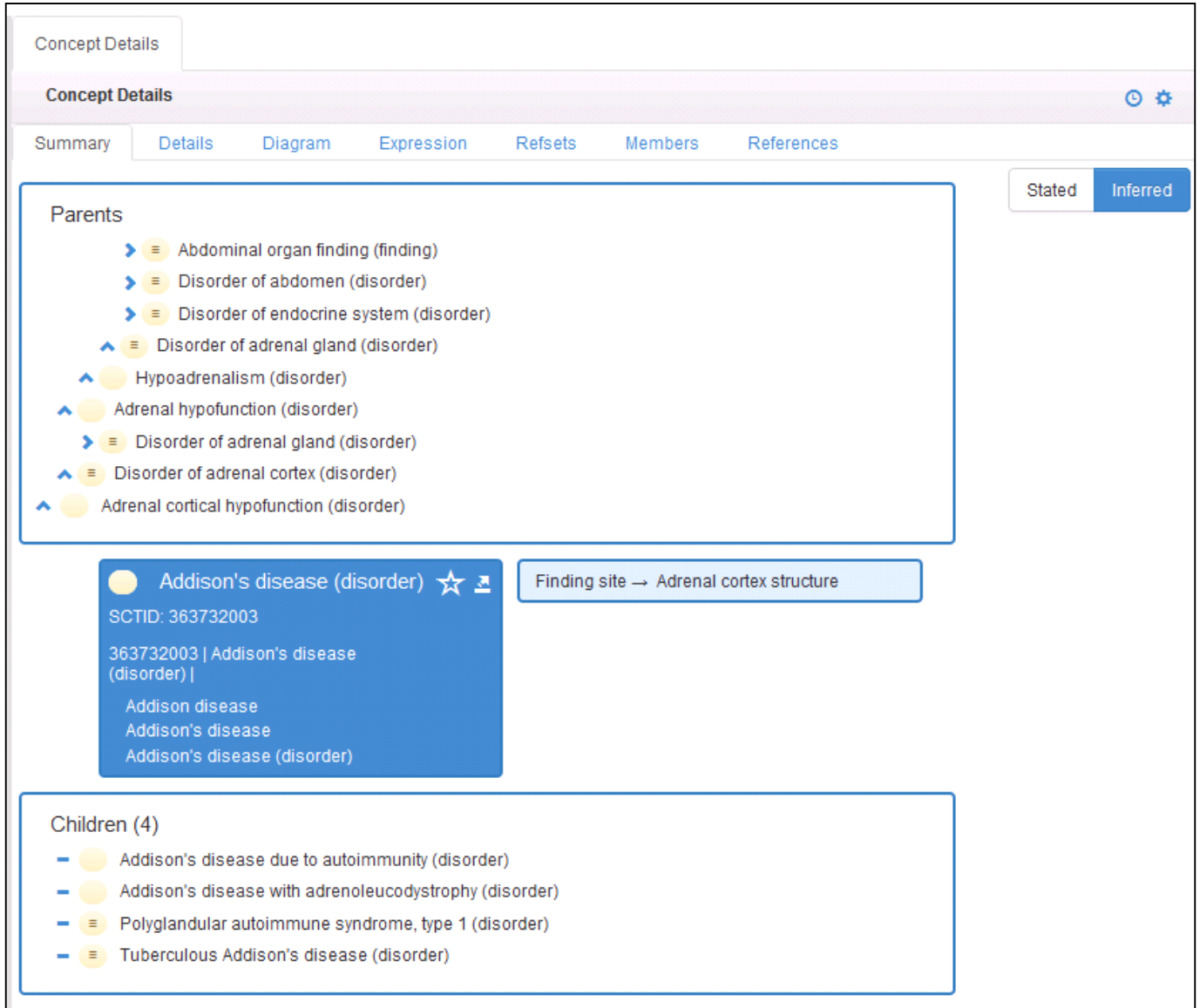
MeSH Tree Structures

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] +

[...]

SNOMED CT


Concept Details
Concept Details
Summary
Details
Diagram
Expression
Refsets
Members
References
Stated
Inferred
Parents
Abdominal organ finding (finding)
Disorder of abdomen (disorder)
Disorder of endocrine system (disorder)
Disorder of adrenal gland (disorder)
Hypoadrenalism (disorder)
Adrenal hypofunction (disorder)
Disorder of adrenal gland (disorder)
Disorder of adrenal cortex (disorder)
Adrenal cortical hypofunction (disorder)
Addison's disease (disorder)
SCTID: 363732003
363732003 | Addison's disease (disorder) |
Addison disease
Addison's disease
Addison's disease (disorder)
Finding site → Adrenal cortex structure
Children (4)
Addison's disease due to autoimmunity (disorder)
Addison's disease with adrenoleucodystrophy (disorder)
Polyglandular autoimmune syndrome, type 1 (disorder)
Tuberculous Addison's disease (disorder)

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

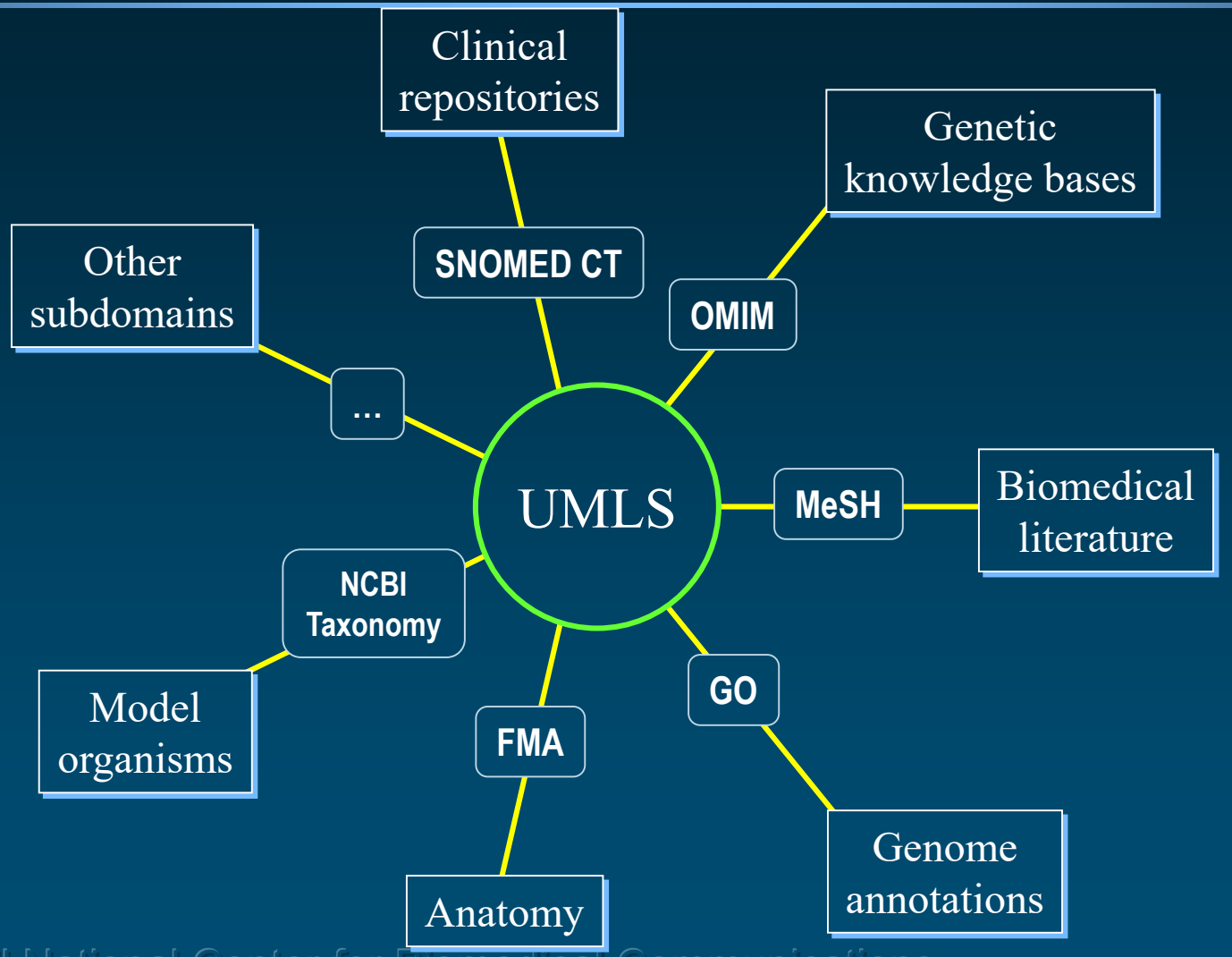
C10AA **HMG CoA reductase inhibitors**

ATC code	Name	DDD	U	Adm.R
C10AA05	atorvastatin	20	mg	O

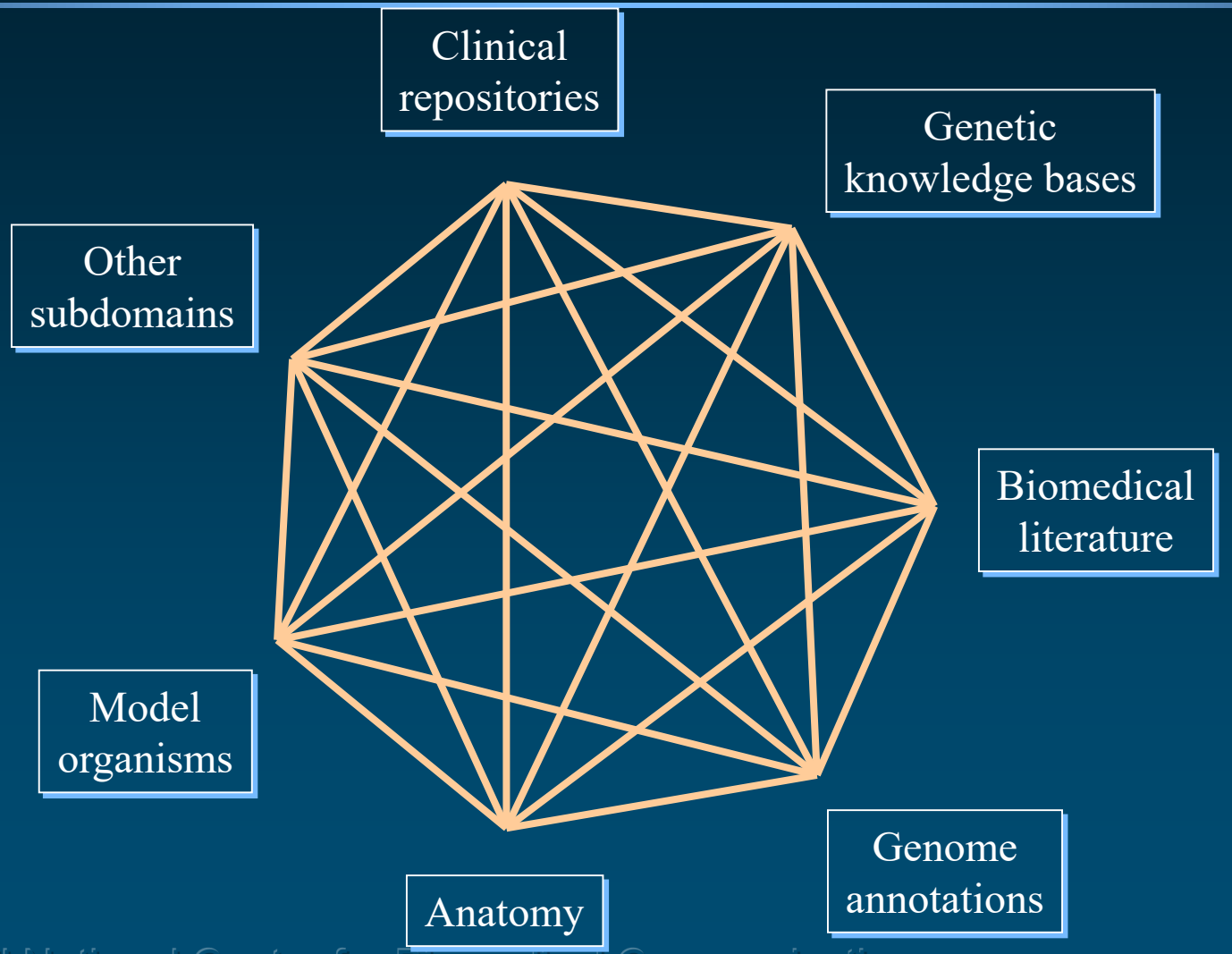
Degrees of semantic interoperability

- 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

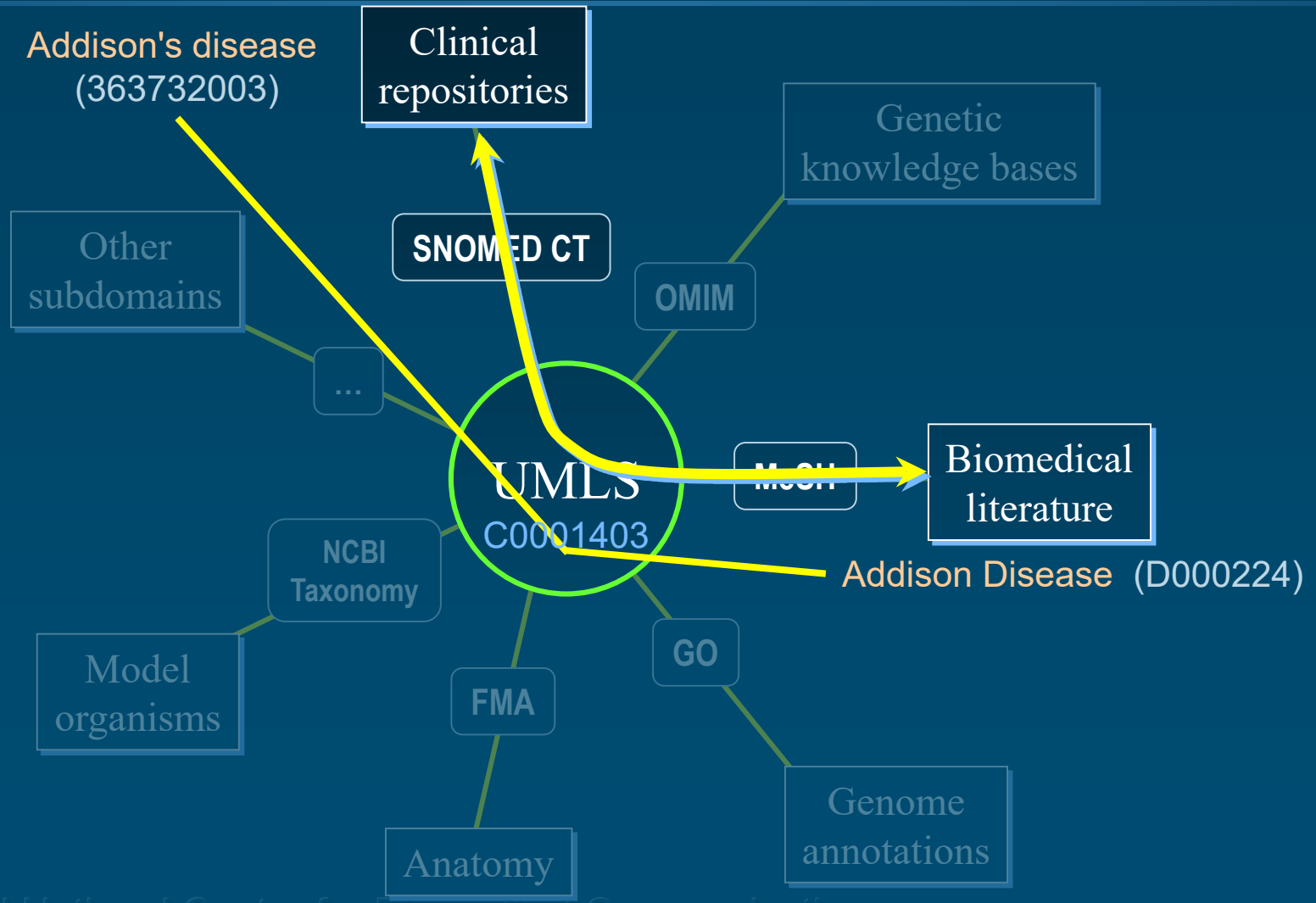
Integrating datasets



Point-to-point mappings are impractical



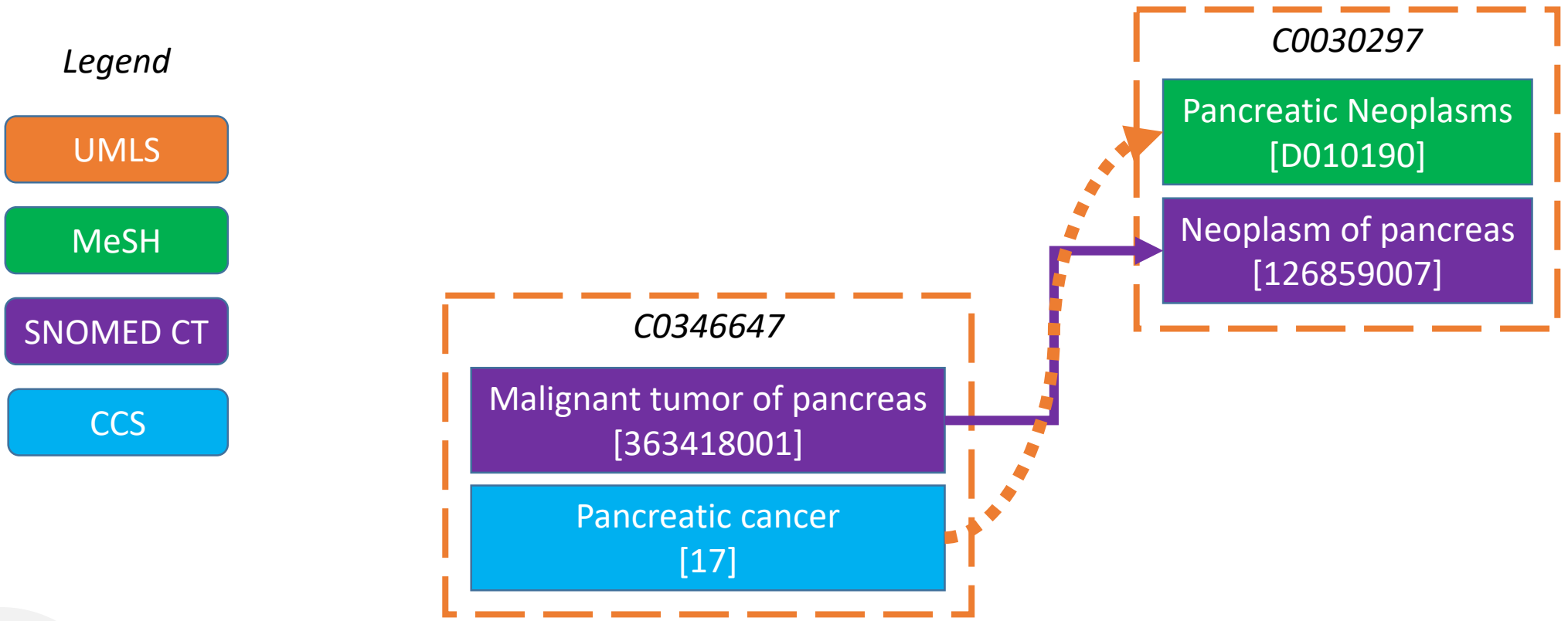
Integration through a reference (e.g., UMLS)



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



Analyzing Opioid Prescriptions in Medicare

Medicare Part D dataset

https://www.resdac.org/cms-data/files/pde/data-documentation

- 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

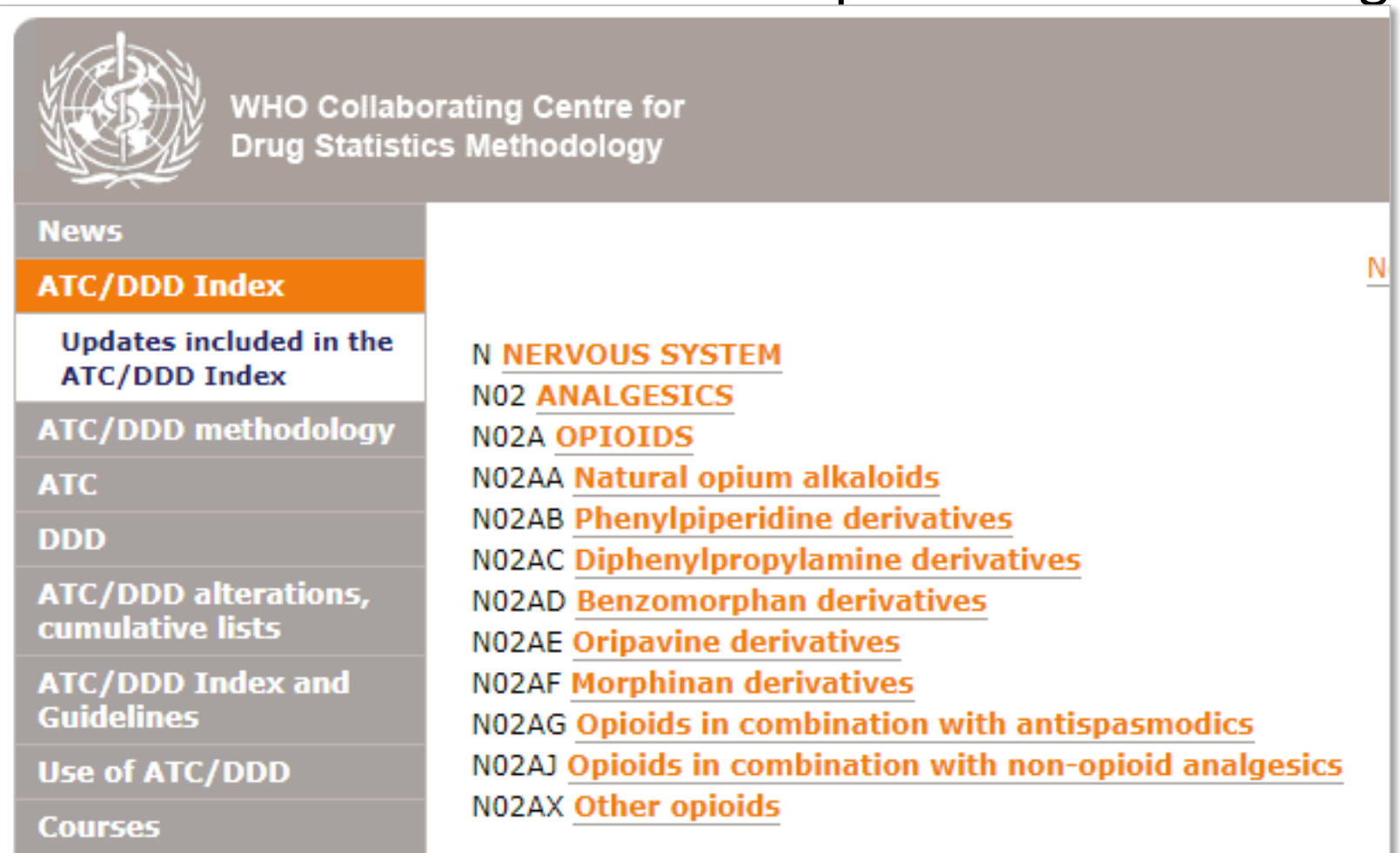
- Identify prescriptions corresponding to opioids in the Medicare part D dataset
- For each opioid drug, calculate the trend of dispensation over time (“number of prescriptions”)
- For all opioids, calculate the trend of total (or daily) dose dispensed in oral morphine milligram equivalents

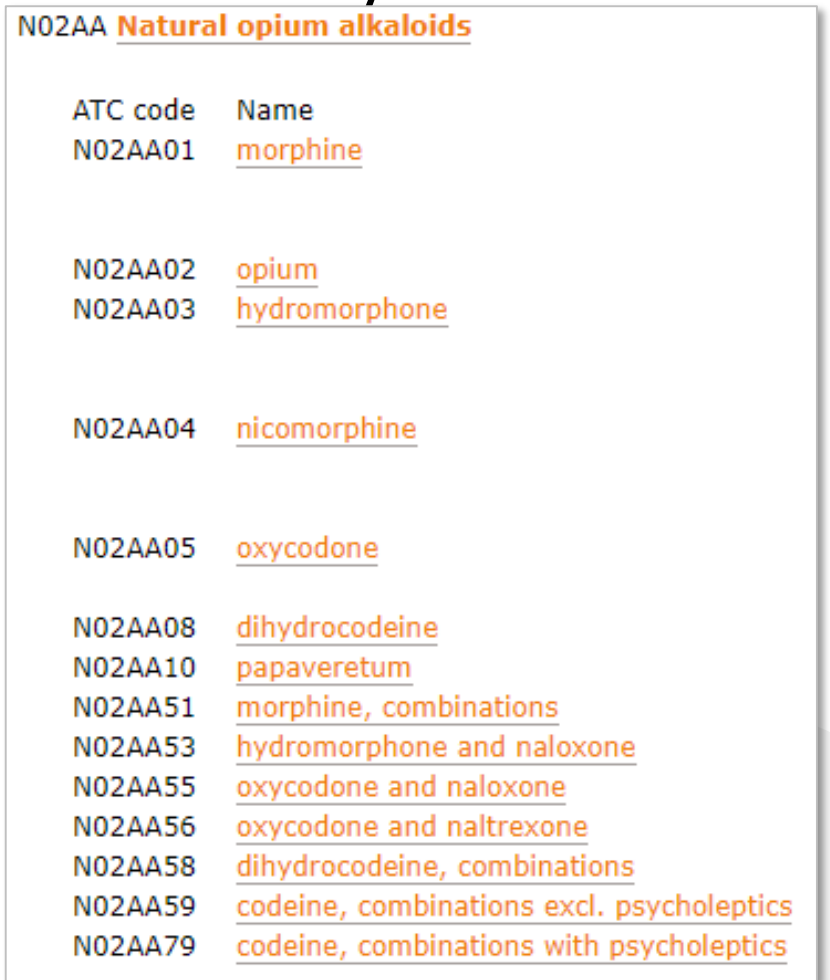
Analysis of opioid prescriptions – How to?

- 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





Linkages among drug entities

N NERVOUS SYSTEM
N02 ANALGESICS
N02A OPIOIDS
N02AA Natural opium alkaloids

ATC code	Name	DDD	U	Adm.R	Note
N02AA05	oxycodone	75	mg	O	
		30	mg	P	

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]

59011041010

59011-410-10

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

https://www.cms.gov/Medicare/Prescription-Drug-Coverage/PrescriptionDrugCovContra/Downloads/Opioid-Morphine-EQ-Conversion-Factors-Aug-2017.pdf

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

Opioid Oral Morphine Milligram Equivalent (MME) Conversion Factors[1,2]

Type of Opioid (strength units)	MME Conversion Factor
Buprenorphine film/tablet[3] (mg)	30
Buprenorphine patch[4] (mcg/hr)	12.6
Buprenorphine film (mcg)	0.03
Butorphanol (mg)	7
Codeine (mg)	0.15
Dihydrocodeine (mg)	0.25
Fentanyl buccal or SL tablets, or lozenge/troche[5] (mcg)	0.13
Fentanyl film or oral spray[6] (mcg)	0.18
Fentanyl nasal spray[7] (mcg)	0.16
Fentanyl patch[8] (mcg)	7.2
Hydrocodone (mg)	1
Hydromorphone (mg)	4
Levorphanol tartrate (mg)	11
Meperidine hydrochloride (mg)	0.1
Methadone[9] (mg)	3
>0, <= 20	4
>20, <=40	8
>40, <=60	10
>60	12
Morphine (mg)	1
Opium (mg)	1
Oxycodone (mg)	1.5
Oxymorphone (mg)	3
Pentazocine (mg)	0.37
Tapentadol [10] (mg)	0.4
Tramadol (mg)	0.1

Example #1

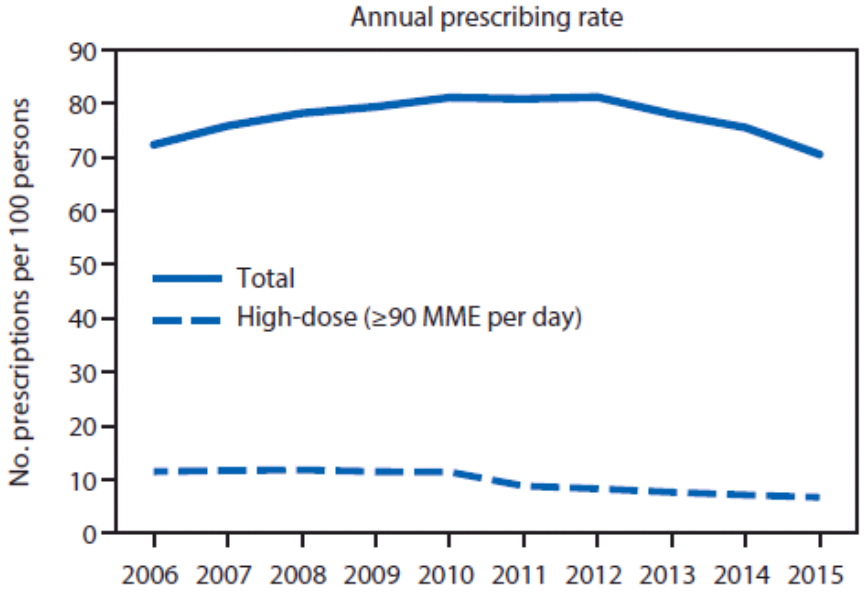
- Drug: NDC = 59011041010
  - *12 HR OxyCONTIN 10 MG Extended Release Oral Tablet [1049504]*
  - Ingredient: Oxycodone
  - Strength: 10 MG
- Dispensation information
  - Quantity Dispensed: 20
  - Days Supply: 10
- MME conversion factor: 1.5
- MME/day:

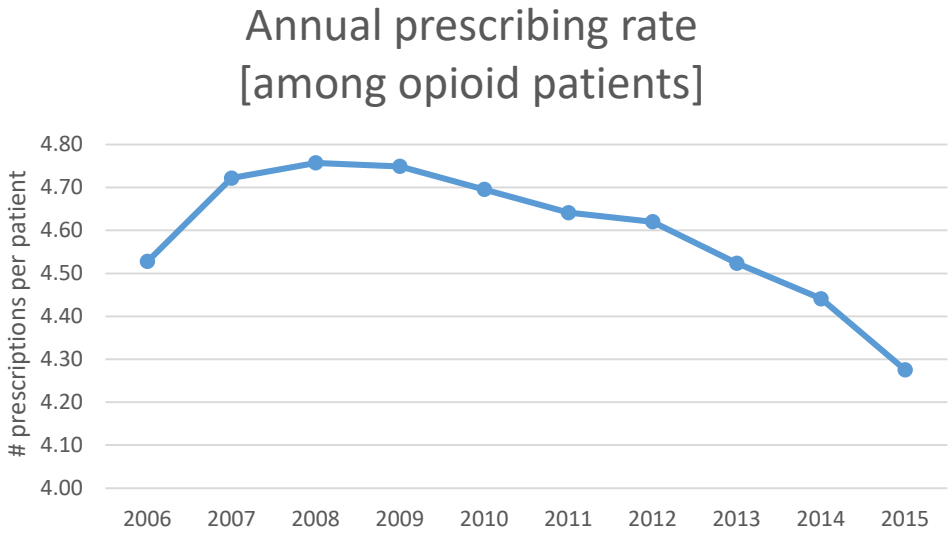
$$MME(mg) = 10 * \frac{20}{10} * 1.5 = 30\ mg$$

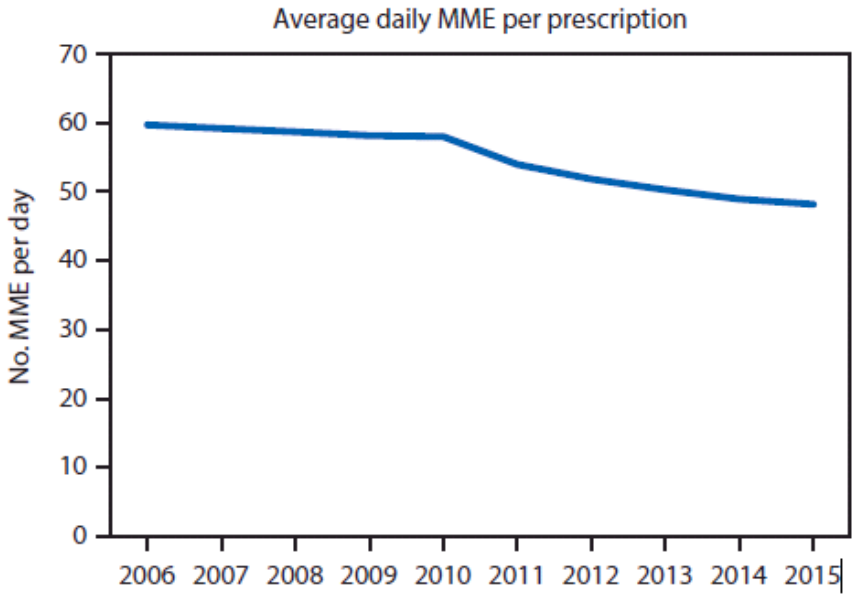
US Department of Health and Human Services/Centers for Disease Control and Prevention

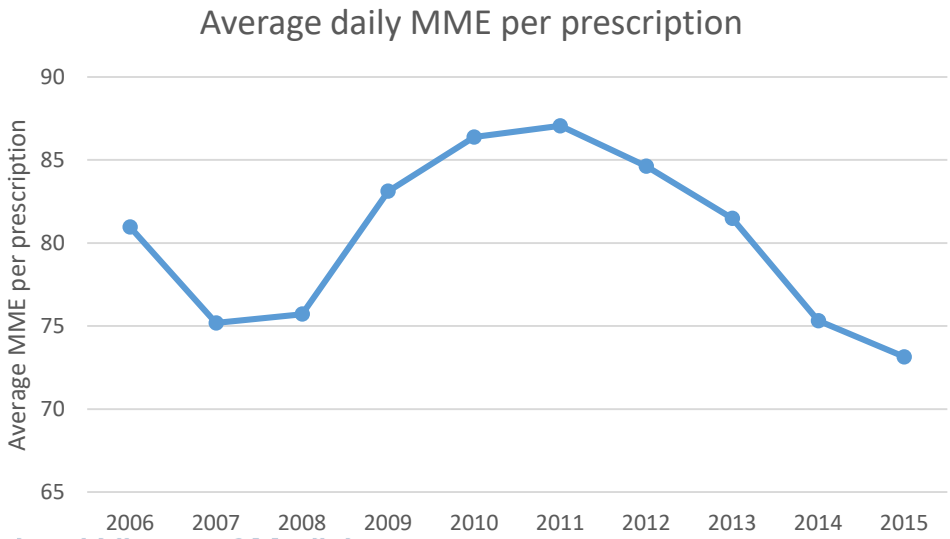
MMWR / July 7, 2017 / Vol. 66 / No. 26

Medicare data










olivier.bodenreider@nih.gov
NIH
U.S. National Library of Medicine