SNOMED CT
Ontologies for Clinical Value Symposium

London, UK
April 12, 2018

The role of terminologies in health data analytics through common data models

Olivier Bodenreider

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

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

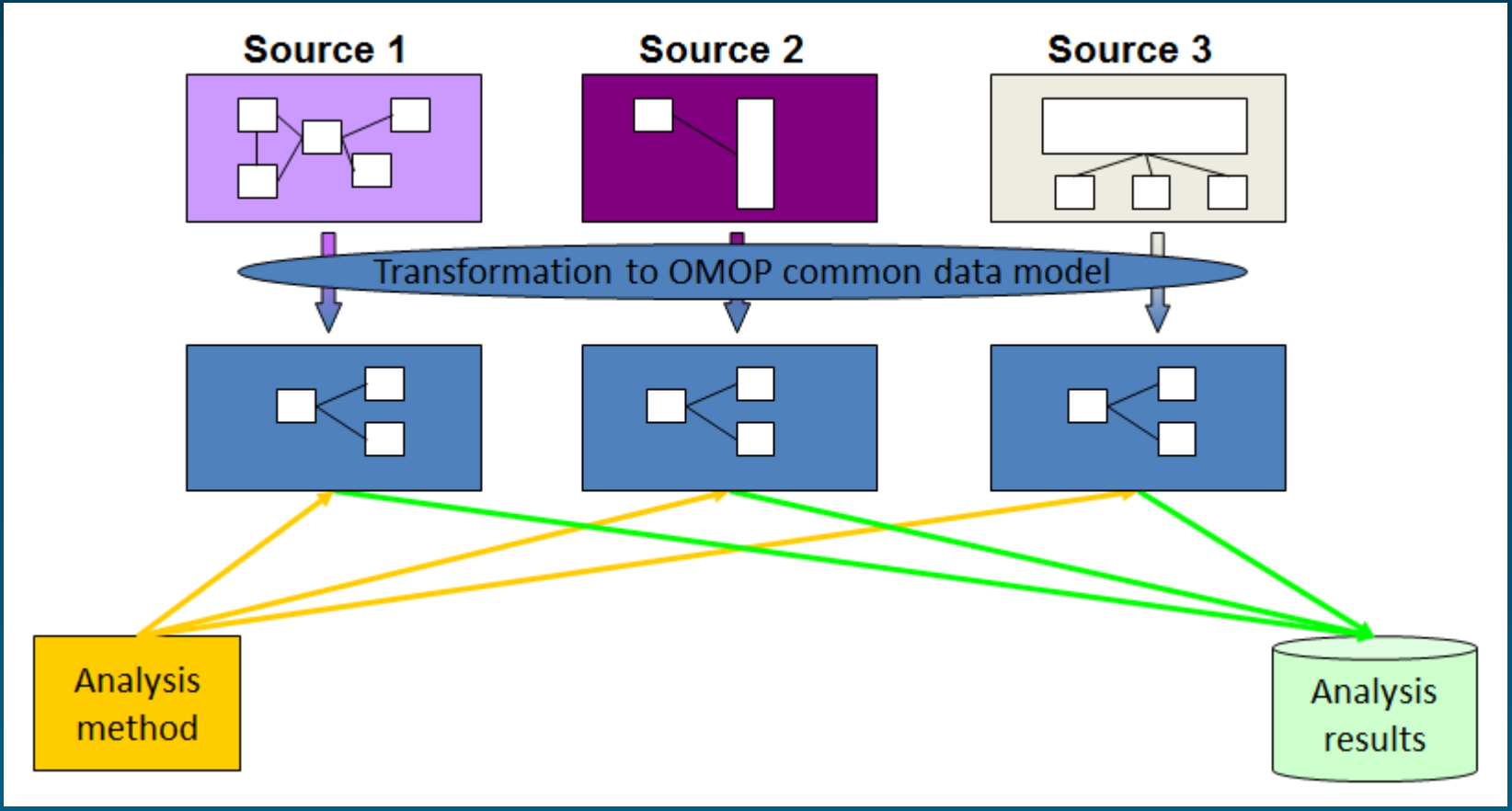
- The context of health data analytics
  - Data models
  - Terminology integration
- Observational Health Data Sciences and Informatics (OHDSI)

“Common” data models

- OMOP
- i2b2
- PCORnet
- Sentinel
- CDISC

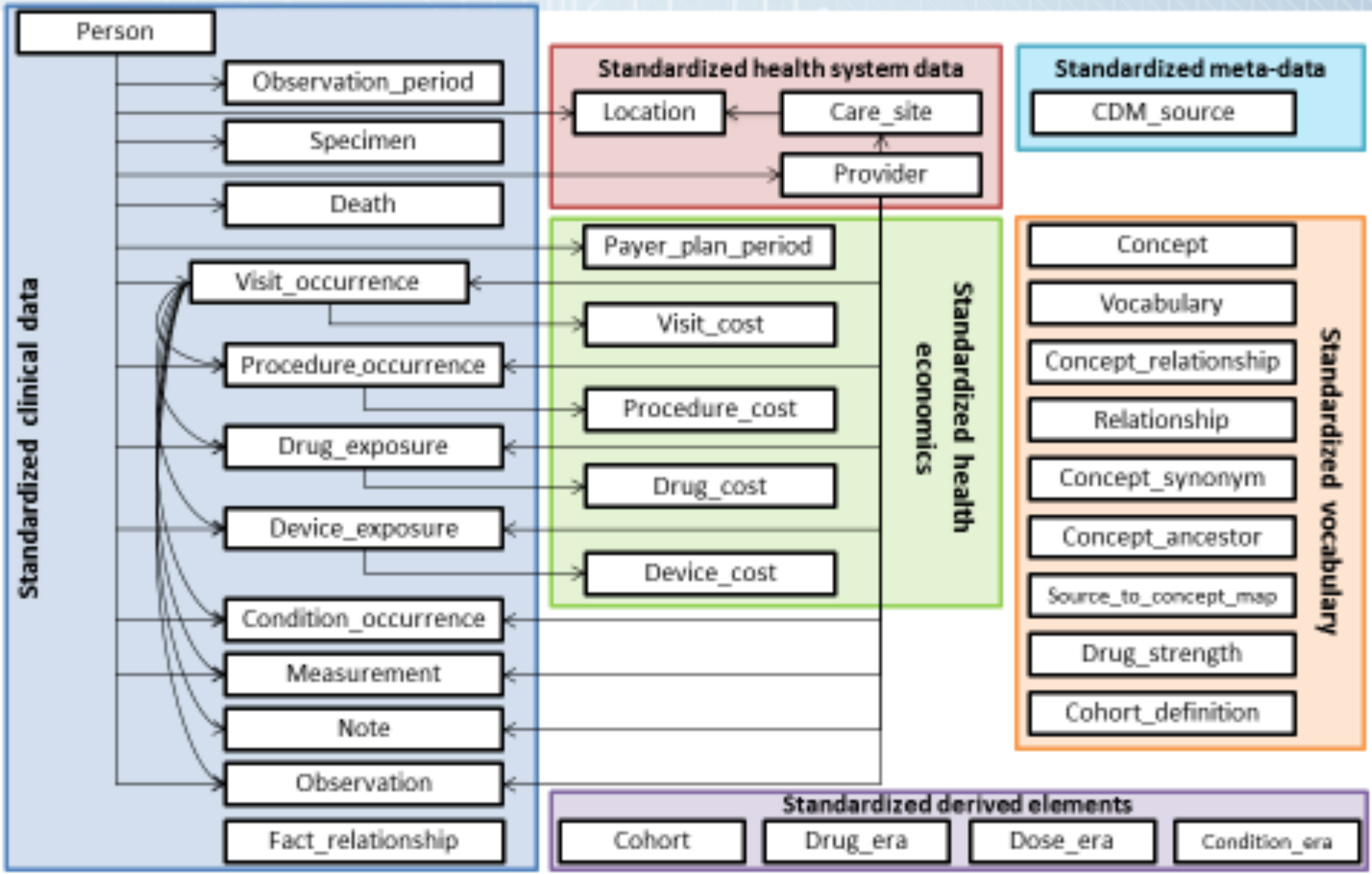
OMOP

- OMOP – Observational Medical Outcomes Partnership



Courtesy of Christian Reich

OMOP Common Data Model



http://www.ohdsi.org/web/wiki/doku.php?id=documentation:cdm

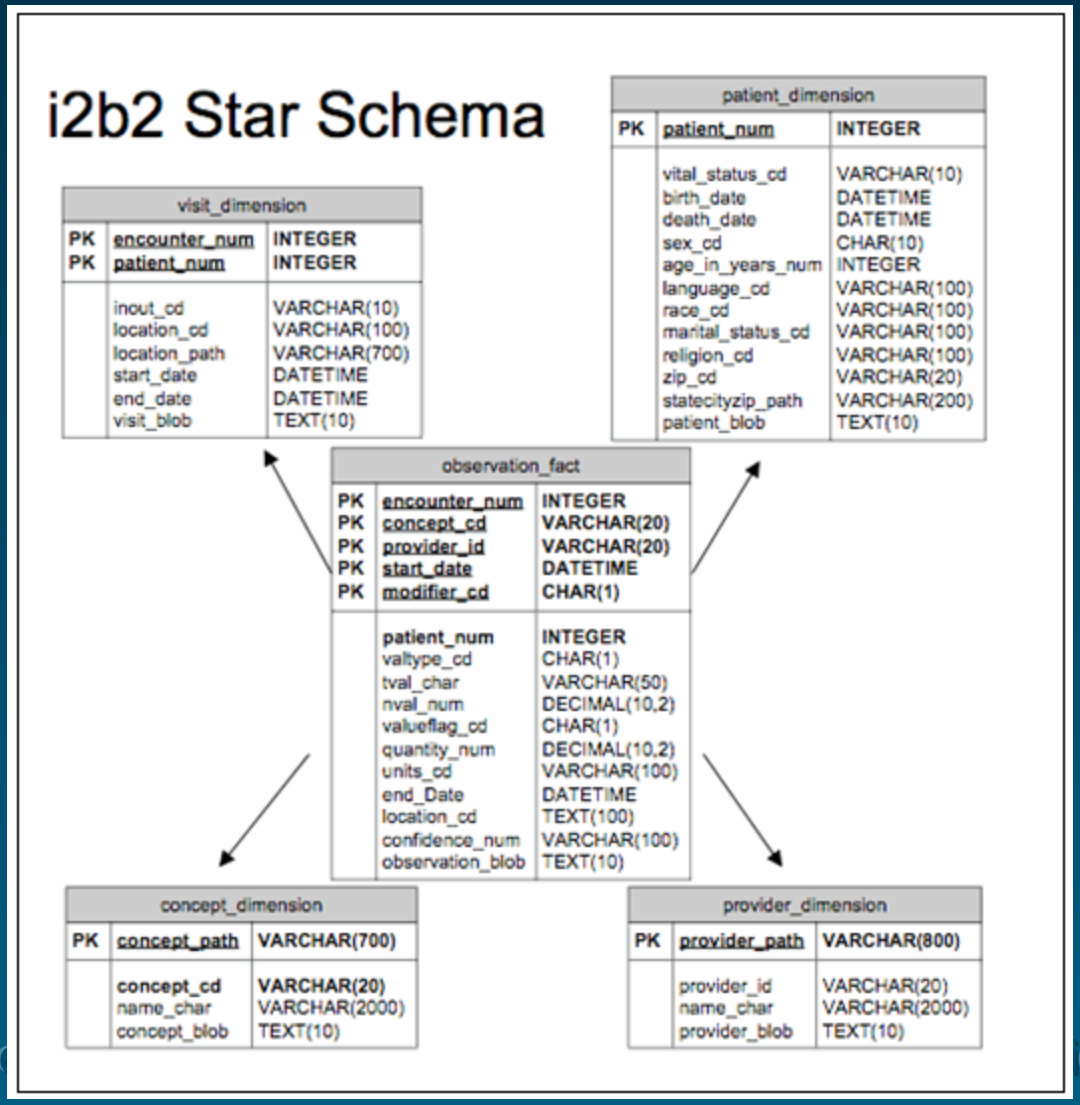
- Standardized Clinical Data Tables
  - PERSON
  - OBSERVATION_PERIOD
  - SPECIMEN
  - DEATH
  - VISIT_OCCURRENCE
  - PROCEDURE_OCCURRENCE
  - DRUG_EXPOSURE
  - DEVICE_EXPOSURE
  - CONDITION_OCCURRENCE
  - MEASUREMENT
  - NOTE
  - NOTE_NLP (V5 5.2)
  - OBSERVATION
  - FACT_RELATIONSHIP
- Standardized 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

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

- 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

New to v3.0

DEMOGRAPHIC
PATID
BIRTH_DATE
BIRTH_TIME
SEX
HISPANIC
RACE
BIOBANK_FLAG

Fundamental basis

ENROLLMENT
PATID
ENR_START_DATE
ENR_END_DATE
CHART
ENR_BASIS

DISPENSING
DISPENSINGID
PATID
PRESCRIBINGID (optional)
DISPENSE_DATE
NDC
DISPENSE_SUP
DISPENSE_AMT

DEATH
PATID
DEATH_DATE
DEATH_DATE_IMPUTE
DEATH_SOURCE
DEATH_MATCH_CONFIDENCE

DEATH_CONDITION
PATID
DEATH_CAUSE
DEATH_CAUSE_CODE
DEATH_CAUSE_TYPE
DEATH_CAUSE_SOURCE
DEATH_CAUSE_CONFIDENCE

Data captured from processes associated with healthcare delivery

VITAL
VITALID
PATID
ENCOUNTERID (optional)
MEASURE_DATE
MEASURE_TIME
VITAL_SOURCE
HT
WT
DIASTOLIC
SYSTOLIC
ORIGINAL_BMI
BP_POSITION
SMOKING
TOBACCO
TOBACCO_TYPE

CONDITION
CONDITIONID
PATID
ENCOUNTERID (optional)
REPORT_DATE
RESOLVE_DATE
ONSET_DATE
CONDITION_STATUS
CONDITION
CONDITION_TYPE
CONDITION_SOURCE

PRO_CM
PRO_CM_ID
PATID
ENCOUNTERID (optional)
PRO_ITEM
PRO_LOINC
PRO_DATE
PRO_TIME
PRO_RESPONSE
PRO_METHOD
PRO_MODE
PRO_CAT

Data captured within multiple contexts: healthcare delivery, registry activity, or directly from patients

ENCOUNTER
ENCOUNTERID
PATID
ADMIT_DATE
ADMIT_TIME
DISCHARGE_DATE
DISCHARGE_TIME
PROVIDERID
FACILITY_LOCATION
ENC_TYPE
FACILITYID
DISCHARGE_DISPOSITION
DISCHARGE_STATUS
DRG
DRG_TYPE
ADMITTING_SOURCE

DIAGNOSIS
DIAGNOSISID
PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
DX
DX_TYPE
DX_SOURCE
PDX

PROCEDURES
PROCEDURESID
PATID
ENCOUNTERID
ENC_TYPE (replicated)
ADMIT_DATE (replicated)
PROVIDERID (replicated)
PX_DATE
PX
PX_TYPE
PX_SOURCE

LAB_RESULT_CM
LAB_RESULT_CM_ID
PATID
ENCOUNTERID (optional)
LAB_NAME
SPECIMEN_SOURCE
LAB_LOINC
PRIORITY
RESULT_LOC
LAB_PX
LAB_PX_TYPE
LAB_ORDER_DATE
SPECIMEN_DATE
SPECIMEN_TIME
RESULT_DATE
RESULT_TIME
RESULT_QUAL
RESULT_NUM
RESULT_MODIFIER
RESULT_UNIT
NORM_RANGE_LOW
NORM_MODIFIER_LOW
NORM_RANGE_HIGH
NORM_MODIFIER_HIGH
ABN_IND

PRESCRIBING
PRESCRIBINGID
PATID
ENCOUNTERID (optional)
RX_PROVIDERID
RX_ORDER_DATE
RX_ORDER_TIME
RX_START_DATE
RX_END_DATE
RX_QUANTITY
RX_REFILLS
RX_DAYS_SUPPLY
RX_FREQUENCY
RX_BASIS
RXNORM_CUI

Data captured from healthcare delivery, direct encounter basis

PCORNET_TRIAL
PATID
TRIALID
PARTICIPANTID
TRIAL_SITEID
TRIAL_ENROLL_DATE
TRIAL_END_DATE
TRIAL_WITHDRAW_DATE
TRIAL_INVITE_CODE

Associations with PCORnet clinical trials

HARVEST
NETWORKID
NETWORK_NAME
DATAMARTID
DATAMART_NAME
DATAMART_PLATFORM
CDM_VERSION
DATAMART_CLAIMS
DATAMART_EHR
BIRTH_DATE_MGMT
ENR_START_DATE_MGMT
ENR_END_DATE_MGMT
ADMIT_DATE_MGMT
DISCHARGE_DATE_MGMT
PX_DATE_MGMT
RX_ORDER_DATE_MGMT
RX_START_DATE_MGMT
RX_END_DATE_MGMT
DISPENSE_DATE_MGMT
LAB_ORDER_DATE_MGMT
SPECIMEN_DATE_MGMT
RESULT_DATE_MGMT
MEASURE_DATE_MGMT
ONSET_DATE_MGMT
REPORT_DATE_MGMT
RESOLVE_DATE_MGMT
PRO_DATE_MGMT
REFRESH_DEMOGRAPHIC_DATE
REFRESH_ENROLLMENT_DATE
REFRESH_ENCOUNTER_DATE
REFRESH_DIAGNOSIS_DATE
REFRESH_PROCEDURES_DATE
REFRESH_VITAL_DATE
REFRESH_DISPENSING_DATE
REFRESH_LAB_RESULT_CM_DATE
REFRESH_CONDITION_DATE
REFRESH_PRO_CM_DATE
REFRESH_PRESCRIBING_DATE
REFRESH_PCORNET_TRIAL_DATE
REFRESH_DEATH_DATE
REFRESH_DEATH_CAUSE_DATE

Process-related data

Bold font indicates fields that cannot be null due to primary key definitions or record-level constraints.

https://www.fda.gov/Safety/FDAsSentinelInitiative/

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

https://www.fda.gov/Safety/FDAsSentinelInitiative/

Sentinel Common Data Model

List of Tables

Table Name	Source	Description
1. Enrollment	Created by Data Partners using Data Partner data.	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 PatID, 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.
2. Demographic	Created by Data Partners using Data	The SCDM Demographic Table contains one record per PatID with the most recent information on Birth_Date, Sex, Race/Ethnicity, and Zip Code.
3. Dispensing	Created by Data Partners using Data Partner data.	The SCDM Outpatient Pharmacy Dispensing Table contains one record per unique combination of PatID, NDC, and RxDate. Each record represents an outpatient pharmacy dispensing. Rollback transactions and other adjustments should be processed before populating this table.
4.1 Encounter	Created by Data Partners using Data Partner data.	The SCDM Encounter Table contains one record per PatID 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.
4.2 Diagnosis	Created by Data Partners using Data Partner data.	The SCDM Diagnosis Table contains one record per unique combination of PatID, EncounterID, DX, and DX_CodeType. This table should capture all uniquely recorded diagnoses for all encounters.
4.3 Procedure	Created by Data Partners using Data	The SCDM Procedure Table contains one record per unique combination of PatID, EncounterID, P PX_CodeType. This table should capture all uniquely recorded procedures for all encounters.
5.1 Death	Created by Data Partners using Data Partner data.	The SCDM Death Table contains one record per PatID. When legacy data have conflicting report make a local determination as to which to use. There is typically a 1-2 year lag in death registry
5.2 Cause of Death	Created by Data Partners using Data Partner data.	The SCDM Cause of Death Table contains one record per unique combination of PatID and COD. legacy data have conflicting reports, please make a local determination as to which to use. Ther typically a 1-2 year lag in death registry data.
6.1 Laboratory Result	Created by Data Partners using Data Partner data.	The SCDM Laboratory Result Table contains one record per result/entry. Only include resulted la Data Partners are strongly encouraged to review the comprehensive Sentinel Common Data Mod Laboratory Result Table Documentation for details on how to populate each variable.

List of Tables (cont.)

Table Name
6.2 Vital Signs
7. Inpatient Pharmacy
8. Inpatient Transfusion

Role of terminologies

- Normalization
  - Different datasets may be annotated in reference to different terminologies
  - Identify equivalent (or close) concepts across terminologies
- Aggregation
  - Queries are generally formulated at a high-level
  - Terminologies support aggregation
    - Transitive closure
    - Value sets

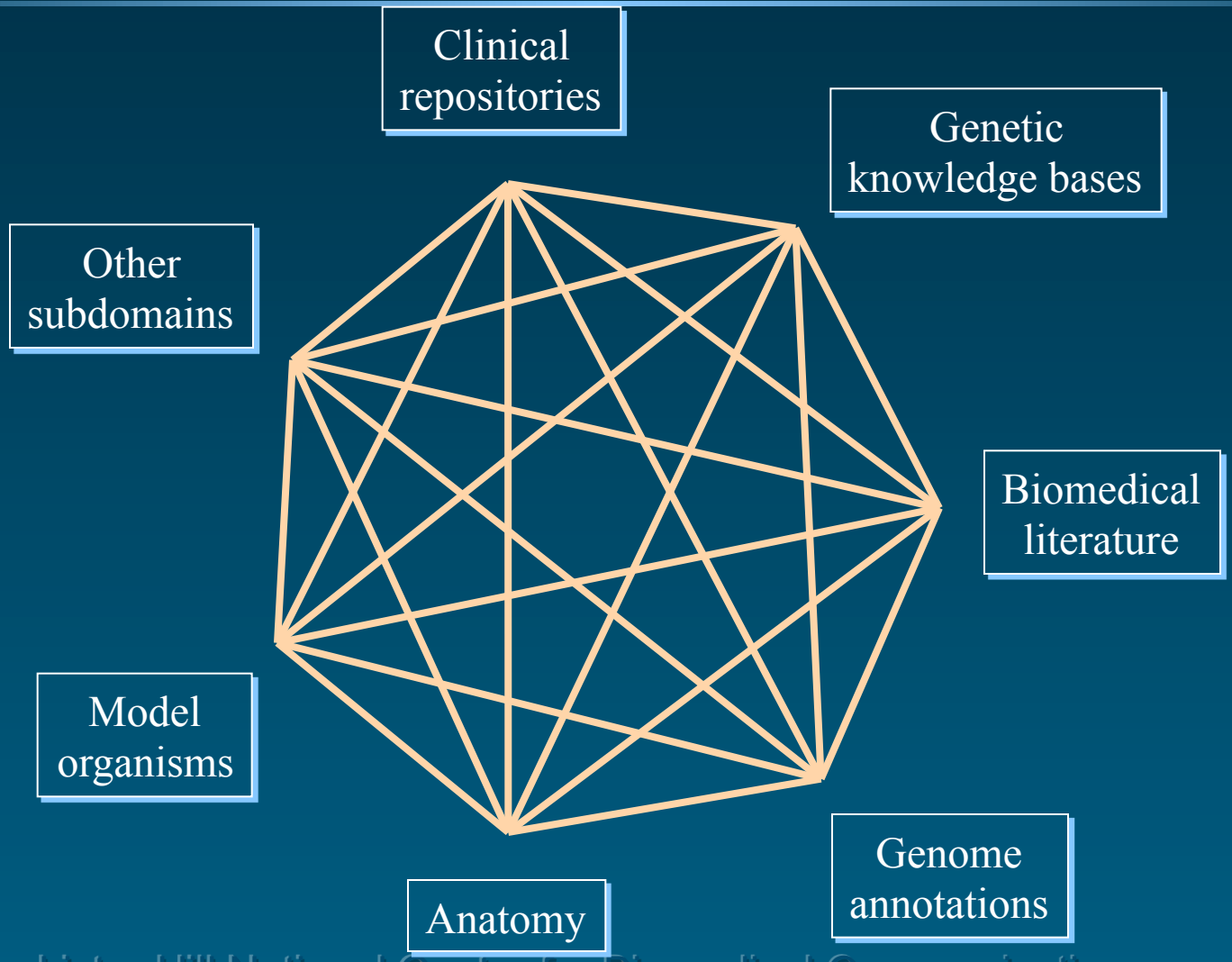
Terminologies used for heath data analytics

- Main clinical terminologies for the Meaningful Use incentive program (clinical documentation; clinical quality measures)
  - SNOMED CT
  - LOINC
  - RxNorm
- Legacy terminologies (billing)
  - [ICD9-CM]; ICD10-CM
  - CPT
- Other terminologies (CDISC)
  - NCI Thesaurus

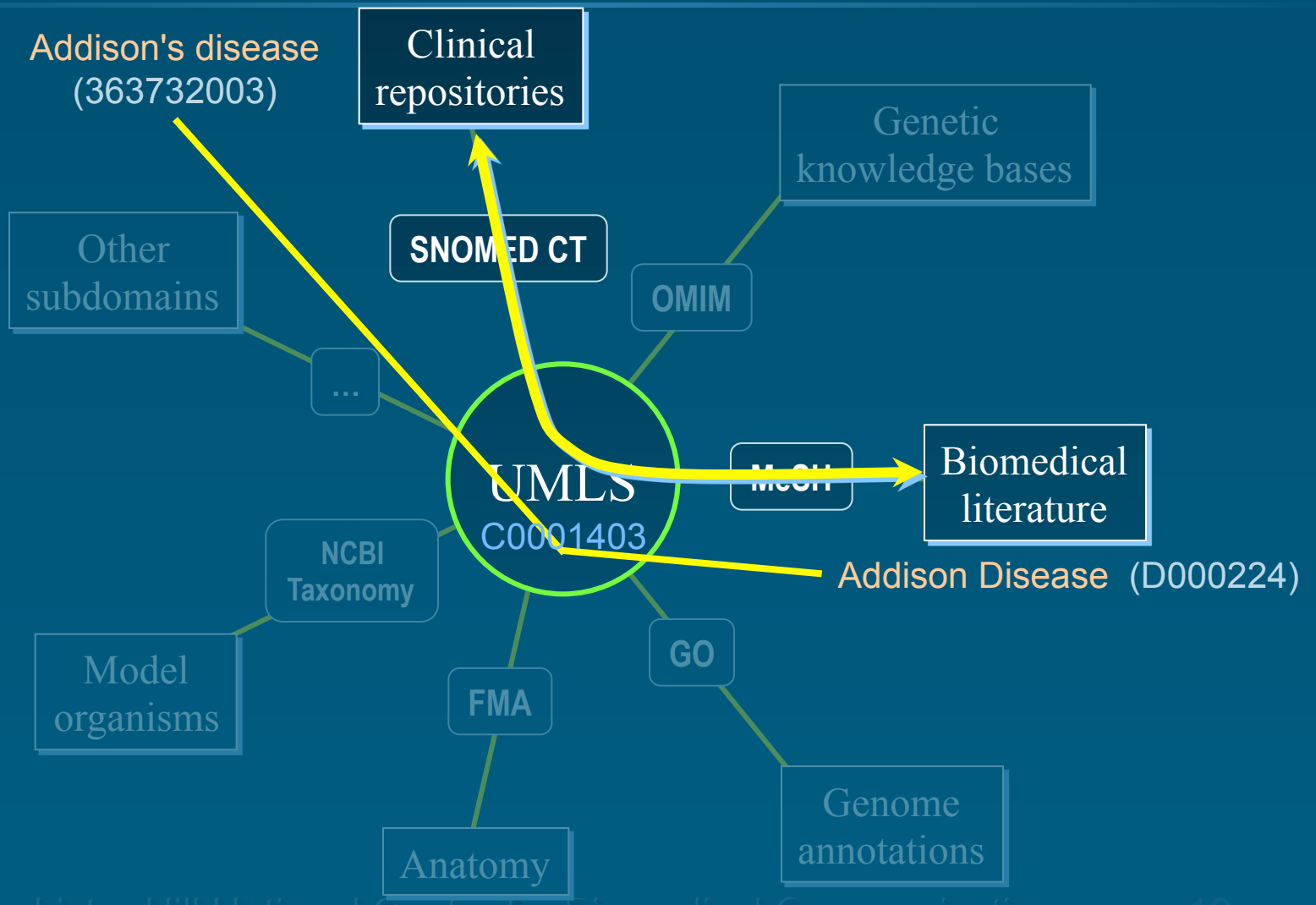
Binding between terminology and information model

- Often involves mapping
- Can happen at various stages along of the integration process
  - ETL – Extract / Transform / Load
  - Query (query translation)
- Issues
  - Integrating heterogeneous datasets
    - Different terminologies
    - Different levels of granularity

Terminology integration



Terminology integration



Observational Health Data Sciences and Informatics (OHDSI)

OHDSI Outline

- From OMOP to OHDSI
- Foundational principles
- OHDSI software, test data and methods
- Use cases and research
  - PNAS paper

From OMOP to OHDSI

- OMOP – Observational Medical Outcomes Partnership
  - Public-private partnership established to inform the appropriate use of observational healthcare databases for studying the effects of medical products (2008-2013)
  - Community of researchers from industry, government, and academia
  - Achievements
    - Conduct methodological research to empirically evaluate the performance of various analytical methods on their ability to identify true associations and avoid false findings
    - Develop tools and capabilities for transforming, characterizing, and analyzing disparate data sources across the health care delivery spectrum
    - Establish a shared resource so that the broader research community can collaboratively advance the science

From OMOP to OHDSI

- OHDSI – Observational Health Data Sciences and Informatics
  - Multi-stakeholder, interdisciplinary collaborative to bring out the value of health data through large-scale analytics
  - International network of researchers and observational health databases with a central coordinating center housed at Columbia University
  - Continues to actively use the OMOP Common Data Model and Standardized Vocabularies
  - Develops open-source solutions [with Greek names]
  - Annual symposium

Foundational principles

- Data standardization through
  - Common data model (OMOP CDM)
  - Standard vocabularies
- Conversion (ETL) of the local clinical data warehouse to the OMOP CDM and standard vocabularies
  - Supported by the WhiteRabbit tool
- Applicable to various types of observational data (EHR, claims)
- Data remain local to a clinical institution
- The same query can be executed at each site and the results aggregated across sites
- Research projects are based on rigorous protocols
- Open-source software

OHDSI software

- ATLAS – unified interface to multiple OHDSI tools
- ATHENA – access to standardized vocabularies
- ACHILLES – database characterization and data quality assessment
- CALYPSO – analytical component for clinical study feasibility assessment
- CIRCE – cohort creation
- HERACLES – cohort-level analysis and visualization
- LAERTES – system for investigating the association of drugs and health (adverse events)
- DRUG EXPOSURE EXPLORER – visualize drug exposures (an experimental deployment using the SynPUF 1% simulated patient data set)

OHDSI methods

- Population-Level Estimation
  - Safety surveillance
  - Comparative effectiveness
- Patient-Level Prediction

- Implemented with open-source tools for large-scale analytics
  - R packages

Examples of network research studies

- Comparison of combination treatment in hypertension
- Comparative effectiveness of alendronate and raloxifene in reducing the risk of hip fracture
- Levetiracetam and risk of angioedema in patients with seizure disorder
- Drug utilization in children
- Characterizing treatment pathways at scale using the OHDSI network

In development

Characterizing treatment pathways at scale using the OHDSI network

CrossMark click for updates

COLLOQUIUM PAPER

Characterizing treatment pathways at scale using the OHDSI network

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

www.pnas.org/cgi/doi/10.1073/pnas.1510502113 PNAS | July 5, 2016 | vol. 113 | no. 27 | 7329–7336

Characterizing treatment pathways at scale using the OHDSI network

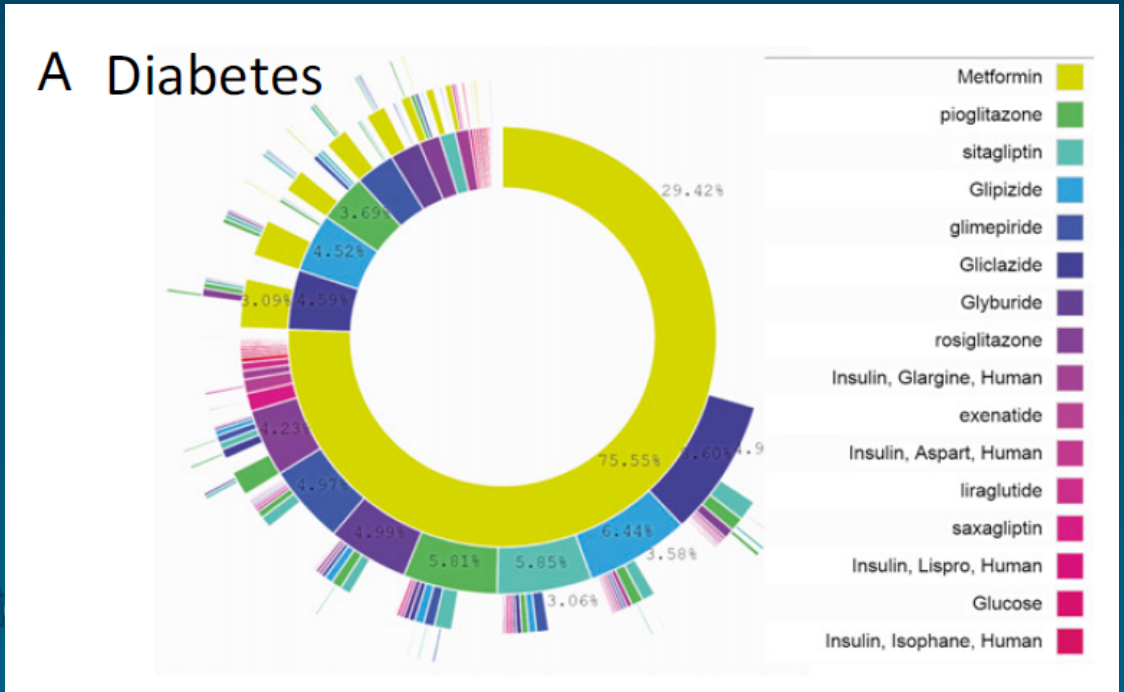
- Objectives: analyze the variability of pharmacological treatment interventions over three years across three diseases (type-2 diabetes mellitus, hypertension, or depression)
- Inclusion criteria: exposure to an antidiabetic, antihypertensive, or antidepressant medication for 3 years, as well as presence of at least one diagnostic code for the corresponding disease
- Exclusion criteria: based on diagnostic data (e.g., exclusion of schizophrenia patients from the depression cohort)

Characterizing treatment pathways at scale using the OHDSI network

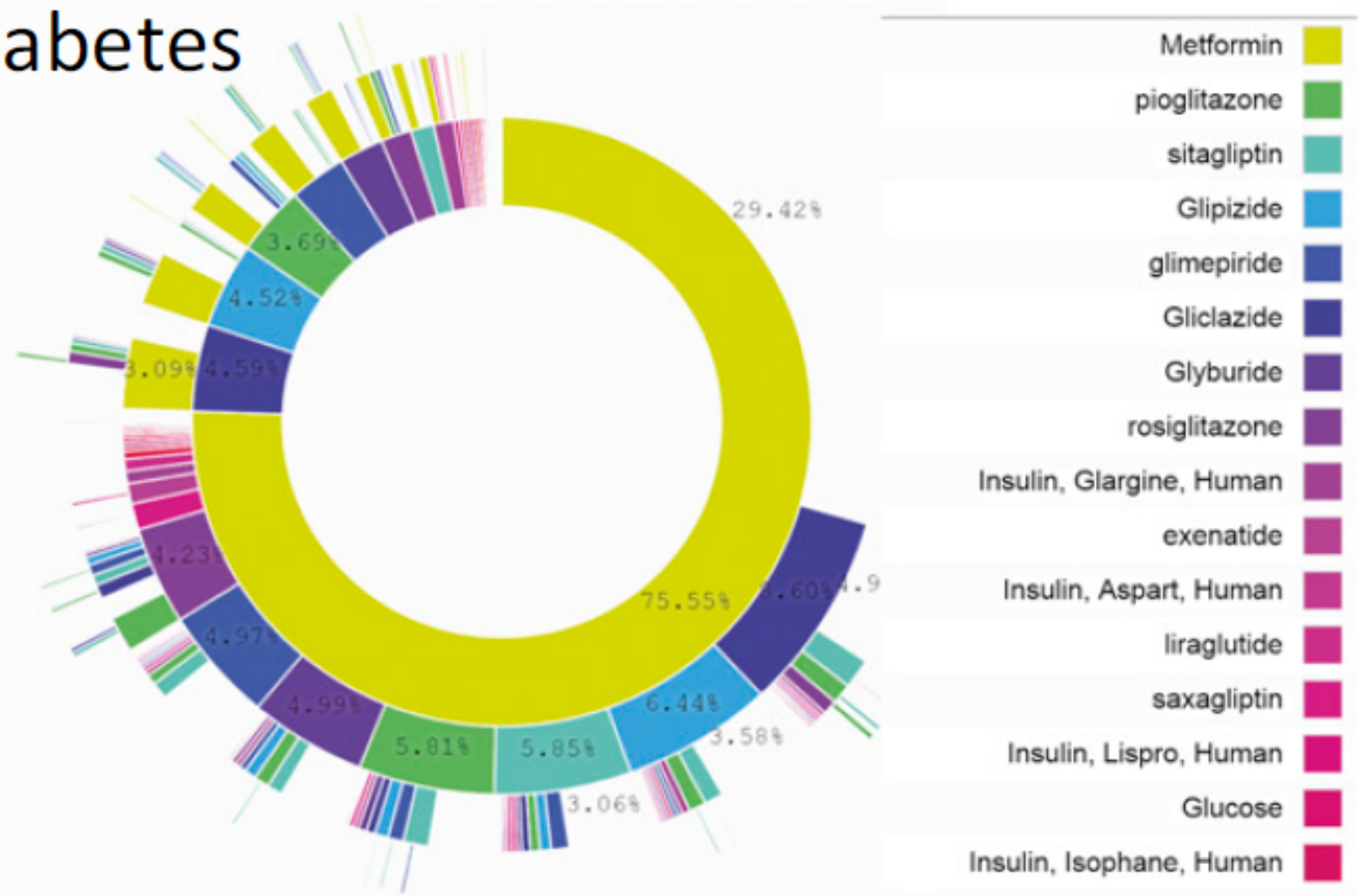
- Materials: 11 datasets representing a total of 255 million patients
  - EHR data (South Korea, U.K., U.S.) 67M
  - Claims data (U.S., Japan) 188M
- Methods: Analyze the sequences of medications that patients were placed on during those 3 years, to reveal patterns and variation in treatment among data sources and diseases

Characterizing treatment pathways at scale using the OHDSI network

- Results
  - Patients with 3 years of uninterrupted therapy
    - 327,110 diabetes patients
    - 1,182,792 hypertension patients
    - 264,841 depression patients
  - Treatment pathways

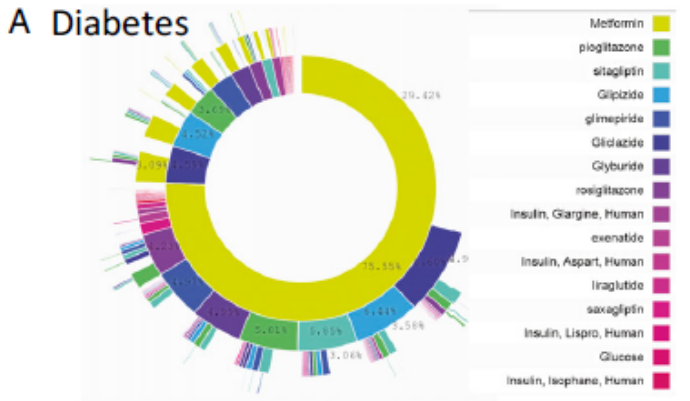


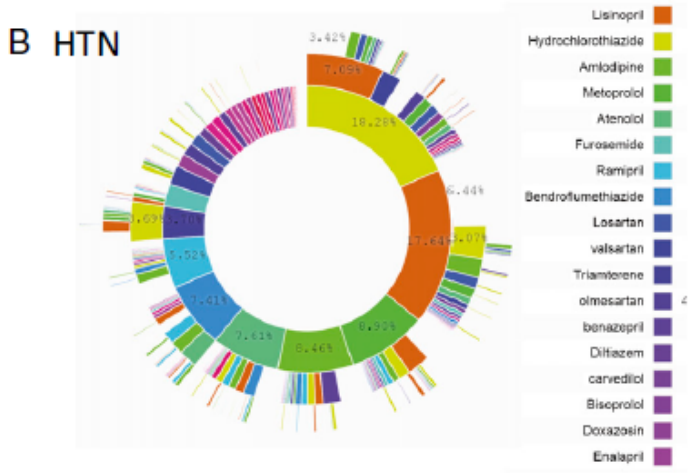
A Diabetes

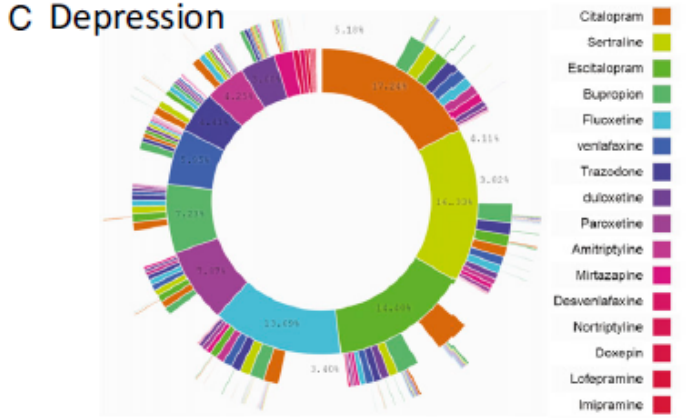

29.42%
75.55%
6.44%
5.85%
5.01%
4.99%
4.97%
4.23%
4.59%
4.52%
3.69%
3.09%
3.58%
3.06%
Metformin
pioglitazone
sitagliptin
Glipizide
glimepiride
Gliclazide
Glyburide
rosiglitazone
Insulin, Glargine, Human
exenatide
Insulin, Aspart, Human
liraglutide
saxagliptin
Insulin, Lispro, Human
Glucose
Insulin, Isophane, Human

Differences across diseases

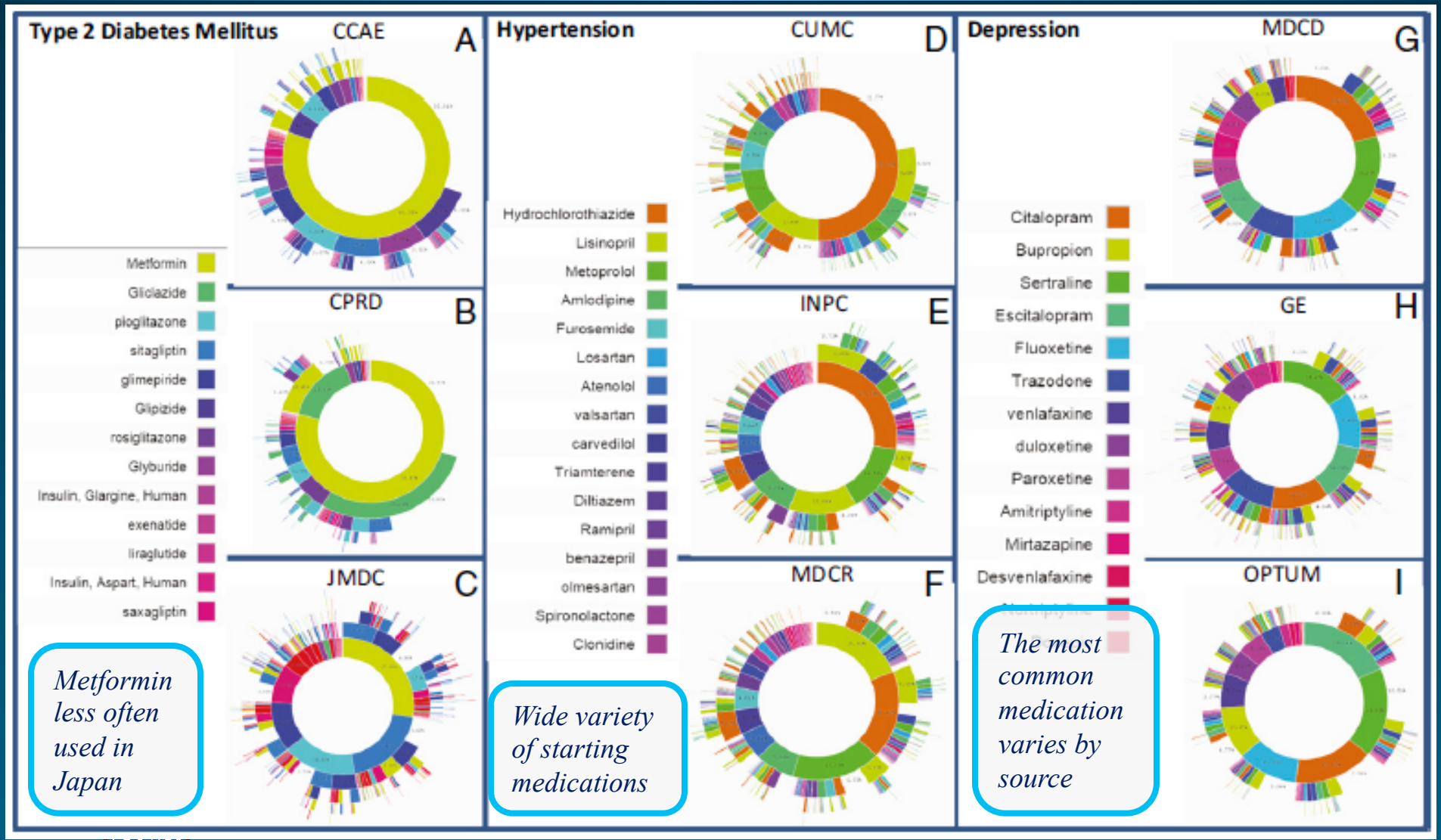
- Diabetes
  - Metformin is the first line of treatment and often the only treatment
- Hypertension
  - Slight predominance of HCTZ, frequently paired with other medications
- Depression
  - Even spread of medications
- Unique treatment pathways (within a cohort)
  - 10% TDM
  - 25% HTN





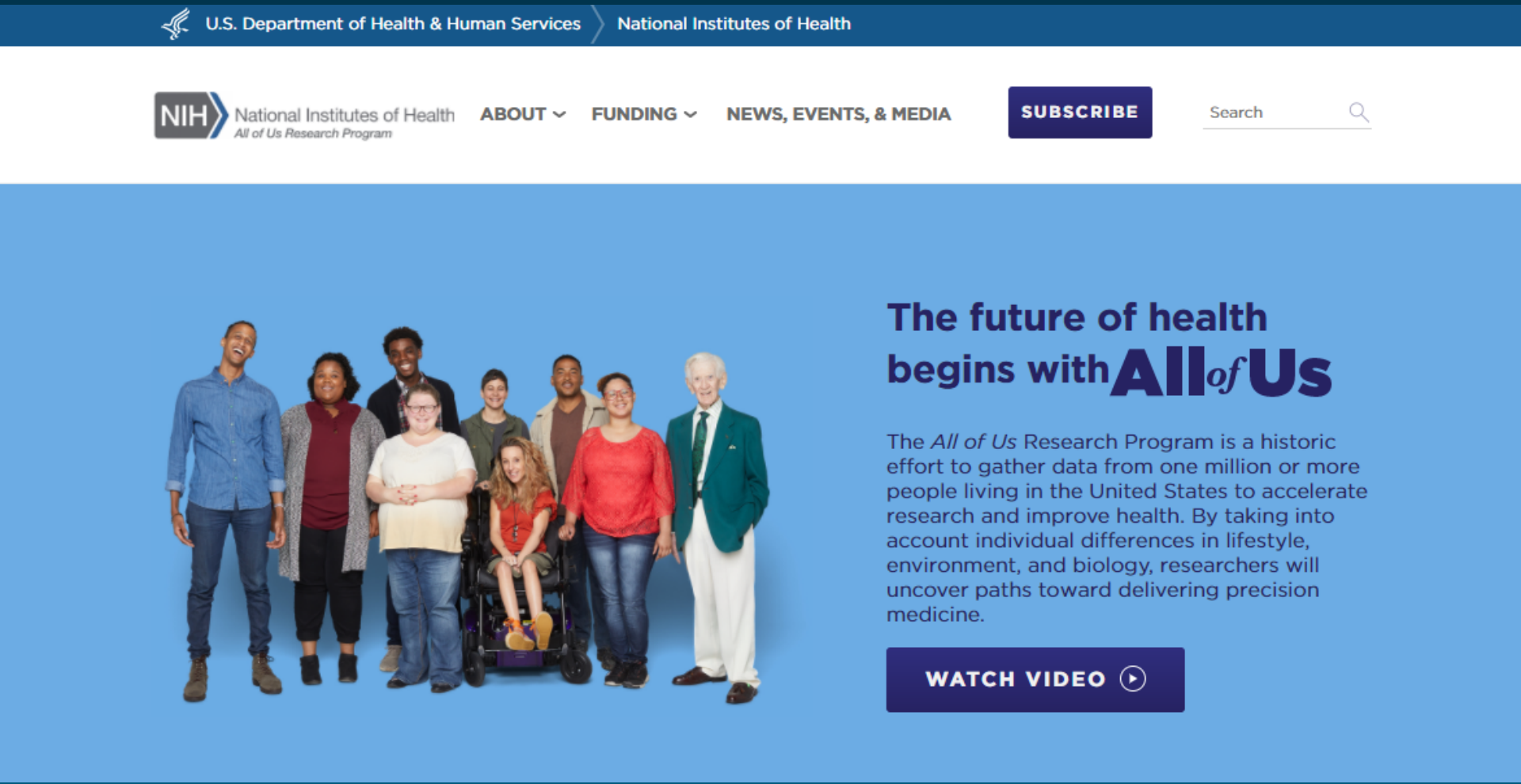


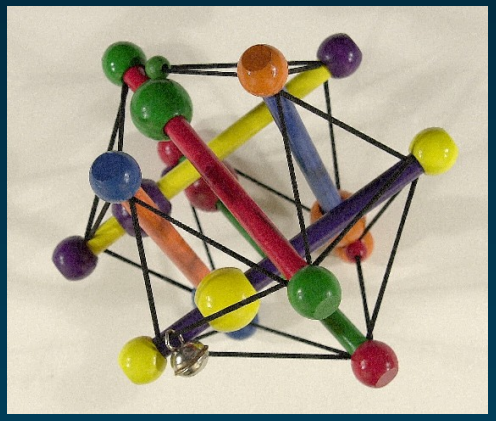
Differences across countries



https://allofus.nih.gov/

All of Us – Precision Medicine Initiative





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

NIH U.S. National Library of Medicine