Interoperability for clinical research

Observational Health Data Sciences and Informatics (OHDSI)

Olivier Bodenreider
Lister Hill National Center for Biomedical Communications
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

◆ The context of clinical data science
  ● Datasets
  ● Data models
  ● Terminologies

◆ Observational Health Data Sciences and Informatics (OHDSI)

◆ Current trends and future directions
The context of clinical data science
The context of clinical data science

- Datasets
- Data models
- Terminologies
Datasets

- Center for Medicare and Medicaid (CMS) data
- MIMIC III
- Clinical data warehouses (academic medical centers, Veterans Administration)
- EHR vendors (GE Centricity; EPIC Clarity; Cerner Health Facts)
- Commercial datasets (OPTUM, Truven)
**Medicare**
Covers: people age 65 or older, people under age 65 with certain disabilities, and people of all ages with End-Stage Renal Disease.
3 parts:
- Part A Hospital Insurance
- Part B Medical Insurance
- Part D Prescription Drug Coverage

**Medicaid**
- Provides health coverage to 69 million Americans, including eligible low-income adults, children, pregnant women, elderly adults and people with disabilities
- Administered by states, according to federal requirements
- Funded jointly by states and the federal government.
Center for Medicare and Medicaid (CMS) data

- Available through the CMS Virtual Research Data Center (VRDC)
  - At a cost
  - Cloud-based environment – data cannot be downloaded

- Longitudinal data available
  - From 1999 for demographics, hospitalization and ambulatory data
  - From 2006 from drug coverage

https://www.resdac.org/cms-data
Example of use of CMS data


**Outpatient beta-blockers and survival from sepsis: Results from a national cohort of Medicare beneficiaries.**

Singer KE¹, Collins CE¹, Flahive JM², Wyman AS¹, Ayturk MD¹, Santry HP³.

**Author information**

**Abstract**

**BACKGROUND:** Elderly Americans suffer increased mortality from sepsis. Given that beta-blockers have been shown to be cardioprotective in critical care, we investigated outpatient beta-blocker prescriptions and mortality among Medicare beneficiaries admitted for sepsis.

**METHODS:** We queried a 5% random sample of Medicare beneficiaries for patients admitted with sepsis. We used in-hospital and outpatient prescription drug claims to compare in-hospital and 30-day mortality based on pre-admission beta-blocker prescription and class of beta-blocker prescribed using univariate tests of comparison and multivariable logistic regression models and another class of medications for control.

**RESULTS:** Outpatient beta-blocker prescription was associated with a statistically significant decrease in in-hospital and 30-day mortality. In multivariable modeling, beta-blocker prescription was associated with 31% decrease in in-hospital mortality and 41% decrease in 30-day mortality. Both cardioselective and non-selective beta-blockers conferred mortality benefit.

**CONCLUSIONS:** Our data suggests that there may be a role for preadmission beta-blockers in reducing sepsis-related mortality.

Copyright © 2017. Published by Elsevier Inc.

PMID: 28666578 DOI: 10.1016/j.amjsurg.2017.06.007
MIMIC III

- Medical Information Mart for Intensive Care
- Freely available to researchers worldwide
- Encompasses a diverse and very large population of ICU patients (~40k)
- Includes demographics, vital signs, laboratory tests, medications, and bedside monitor trends and waveforms
- Contains high temporal resolution data

https://mimic.physionet.org/
Example of use of MIMIC data

Lower short- and long-term mortality associated with overweight and obesity in a large cohort study of adult intensive care unit patients

Swapna Abhyankar, Kira Leishear, Fiona M Callaghan, Dina Demner-Fushman and Clement J McDonald

Received: 2 August 2012 | Accepted: 13 December 2012 | Published: 18 December 2012
Data models

- OMOP
- i2b2
- PCORnet
- Sentinel
- CDISC
OMOP

- 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)

http://www.pcornet.org/
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>Created by Data Partners using Data Partner data.</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, PX, 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/
Terminologies

- 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
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)
**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>
RxNorm - Example

Navigating RxNorm
Drugs

Warfarin [RxCUI = 11289]

IN/MIN
(1)
H Rxs Warfarin

PIN
(2)
S Warfarin Potassium
H Rxs Warfarin Sodium
H Rxs Jantoven

BN
(2)
H Rxs Coumadin
H Rxs Jantoven

SCDC
(11)
S Warfarin Sodium 0.5 MG
S Warfarin Sodium 1 MG
S Warfarin Sodium 10 MG
S Warfarin Sodium 2 MG

SBDC
(19)
H Rxs Warfarin Sodium 1 MG [Coumadin]
H Rxs Warfarin Sodium 1 MG [Jantoven]

SCD/GPCK
(11)
S Warfarin Sodium 0.5 MG Oral Tablet
S Warfarin Sodium 1 MG Oral Tablet
S Warfarin Sodium 10 MG Oral Tablet
S Warfarin Sodium 2 MG Oral Tablet

SBD/BPCK
(19)
H Rxs Coumadin 1 MG Oral Tablet
H Rxs Coumadin 10 MG Oral Tablet
H Rxs Coumadin 2 MG Oral Tablet
H Rxs Coumadin 2 MG/MI Injectable

SCDG
(3)
S Warfarin Injectable Product
S Warfarin Oral Product
S Warfarin Pill

DFG
(3)
Hv Rxs Injectable Product
Hv Rxs Oral Product
Hv Rxs Pill

SBDG
(5)
S Coumadin Injectable Product
S Coumadin Oral Product
S Coumadin Pill
S Jantoven Oral Product
ICD-10 vs. ICD-10-CM

E72 Other disorders of amino-acid metabolism

Excl.: abnormal findings without manifest disease (R79.)

disorders of:
  - aromatic amino-acid metabolism (E70.-)
  - branched-chain amino-acid metabolism (E71.-)
  - fatty-acid metabolism (E72.3)
  - purine and pyrimidine metabolism (E79.-)
gout (M10.-)

E72.0 Disorders of amino-acid transport

Cystine storage disease† (N29.8*)
Cystinosis
Cystinuria
Fanconi(-de Toni)(-Debré) syndrome
Hartnup disease
Lowe syndrome

Excl.: disorders of tryptophan metabolism (E70.8)

E72 Other disorders of amino-acid metabolism

Excludes1: disorders of:
  - aromatic amino-acid metabolism (E70.-)
  - branched-chain amino-acid metabolism (E71.0-E71.2)
  - fatty-acid metabolism (E71.3)
  - purine and pyrimidine metabolism (E79.-)
gout (M1A.-, M10.-)

E72.00 Disorders of amino-acid transport, unspecified

E72.01 Cystinuria

E72.02 Hartnup's disease

E72.03 Lowe's syndrome
  Use additional code for associated glaucoma (H42)

E72.04 Cystinosis
  Fanconi (-de Toni) (-Debré) syndrome with cystinosis

Excludes1: Fanconi (-de Toni) (-Debré) syndrome with...

E72.09 Other disorders of amino-acid transport
  Fanconi (-de Toni) (-Debré) syndrome, unspecified

x6
ICD-10 vs. ICD-10-CM

W58 Bitten or struck by crocodile or alligator

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
Integrating subdomains

- Clinical repositories
- Genetic knowledge bases
- Biomedical literature
- Genome annotations
- Anatomy
- Model organisms
- Other subdomains
Integrating subdomains

Addison's disease (363732003)

Clinical repositories

SNOMED CT

Genetic knowledge bases

UMLS C0001403

Biomedical literature

Addison Disease (D000224)

Other subdomains

NCBI Taxonomy

Model organisms

FMA

GO

Anatomy

Genome annotations
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

http://omop.org
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

[https://www.ohdsi.org/](https://www.ohdsi.org/)
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 test data – SynPUF

◆ 1000-person dataset from the CMS 2008-2010 Data Entrepreneurs’ Synthetic Public Use File (DE-SynPUF)

◆ Converted to the OMOP Common Data Model Version 5

◆ Available as a set of csv files and scripts to load the data into a PostgreSQL database

◆ Useful for testing purposes
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

George Hripcsak, Patrick B. Ryan, Jon D. Duke, Nigam H. Shah, Rae Woong Park, Vojtech Huser, Marc A. Suchard, Martijn J. Schuemie, Frank J. DeFalco, Adler Perotte, Juan M. Banda, Christian G. Reich, Lisa M. Schilling, Michael E. Matheny, Daniella Meeker, Nicole Pratt, and David Madigan

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

Metformin less often used in Japan

Wide variety of starting medications

The most common medication varies by source
Current trends and future directions
ONC interoperability roadmap

The goals are:

- **2015-2017**: Send, receive, find and use priority data domains to improve health care quality and outcomes.
- **2018-2020**: Expand data sources and users in the interoperable health IT ecosystem to improve health and lower costs.
- **2021-2024**: Achieve nationwide interoperability to enable a learning health system, with the person at the center of a system that can continuously improve care, public health, and science through real-time data access.
Learning health system

Best Care at Lower Cost: The Path to Continuously Learning Health Care in America
IOM, May 2013
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
All of Us – Precision Medicine Initiative

The future of health begins with All of Us

The All of Us Research Program is a historic effort to gather data from one million or more people living in the United States to accelerate research and improve health. By taking into account individual differences in lifestyle, environment, and biology, researchers will uncover paths toward delivering precision medicine.

WATCH VIDEO