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

◆ Observational Health Data Sciences and Informatics (OHDSI)
  ● OHDSI in action
    ▪ Analyzing drug treatment pathways for 3 chronic diseases from clinical data warehouse in multiple countries
  ● How do they do that?

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

◆ Current trends and future directions
Observational Health Data Sciences and Informatics (OHDSI)
OHDSI Outline

◆ OHDSI in action
  ● Analyzing drug treatment pathways for 3 chronic diseases from clinical data warehouse in multiple countries

◆ How do they do that?
  ● From OMOP to OHDSI
  ● Foundational principles
  ● OHDSI software, test data and methods
Characterizing treatment pathways at scale using the OHDSI network

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

- 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

- Metformin less often used in Japan
- Wide variety of starting medications
- The most common medication varies by source
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/
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 RabbitInAHat tool
- Applicable to various types of observational data (EHR, claims)
- Data remain local to a clinical institution
- The same query/protocol can be executed at each site and the results aggregated across sites
- Research projects are based on rigorous protocols
- **Open science** (Open-source software, open protocols, open results)
OHDSI (open-source) software

- Tools for mapping data to OHDSI
  - **WhiteRabbit**: profile your source data
  - **RabbitInAHat**: map your source structure to CDM tables and fields
  - **ATHENA**: standardized vocabularies for all CDM domains
  - **Usagi**: map your source codes to CDM vocabulary
  - **ACHILLES**: database characterization and data quality assessment

http://github.com/OHDSI
OHDSI (open-source) software

- **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

◆ Types of methods
  ● Population-Level Estimation
    ■ Safety surveillance
    ■ Comparative effectiveness
  ● Patient-Level Prediction

◆ Effect estimation & calibration
  ● Confidence interval calibration

◆ Implemented with open-source tools for large-scale analytics
  ● R packages
Examples of network research studies

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

◆ OHDSI Europe
  https://www.ohdsi-europe.org/

◆ The Book of OHDSI (Aug. 2019)
  https://ohdsi.github.io/TheBookOfOhdsi/
The context of clinical data science
The context of clinical data science  

Outline

◆ 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. It's administered by states, according to federal requirements, and is funded jointly by states and the federal government.

Medicare
Medicaid

Covering more Americans
Making Americans healthier by preventing illness
Coordinating better care & lowering costs

CMS covers 100 million people...
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

https://github.com/OHDSI/CommonDataModel/wiki
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)
Current version is 4.1

PCORnet Common Data Model v3.0

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

Data captured from healthcare delivery, direct encounter basis.

Bold font indicates fields that cannot be null due to primary key definitions or record-level constraints.
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 with 300 million person-years of high quality, unduplicated, curated data
# Sentinel Common Data Model

## Administrative Data

<table>
<thead>
<tr>
<th>Enrollment</th>
<th>Demographic</th>
<th>Dispensing</th>
<th>Encounter</th>
<th>Diagnosis</th>
<th>Procedure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient ID</td>
<td>Patient ID</td>
<td>Patient ID</td>
<td>Patient ID</td>
<td>Patient ID</td>
<td>Patient ID</td>
</tr>
<tr>
<td>Enrollment Start &amp; End Dates</td>
<td>Birth Date</td>
<td>Dispensing Date</td>
<td>Service Date(s)</td>
<td>Encounter ID</td>
<td>Service Date(s)</td>
</tr>
<tr>
<td>Drug Coverage</td>
<td>Sex</td>
<td>National Drug Code (NDC)</td>
<td>Encounter ID</td>
<td>Encounter Type and Provider</td>
<td>Encounter ID</td>
</tr>
<tr>
<td>Medical Coverage</td>
<td>Zip Code</td>
<td>Days Supply</td>
<td>Encounter Type and Provider</td>
<td>Diagnosis Code &amp; Type</td>
<td>Procedure Code &amp; Type</td>
</tr>
<tr>
<td>Medical Record Availability</td>
<td>Etc.</td>
<td>Amount Dispensed</td>
<td>Facility</td>
<td>Principal Discharge Diagnosis</td>
<td>Etc.</td>
</tr>
</tbody>
</table>

## Registry Data

<table>
<thead>
<tr>
<th>Death</th>
<th>Cause of Death</th>
<th>State Vaccine</th>
<th>Inpatient Data</th>
<th>Mother-Infant Linkage Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient ID</td>
<td>Patient ID</td>
<td>Patient ID</td>
<td>Inpatient Pharmacy</td>
<td>Inpatient Transfusion</td>
</tr>
<tr>
<td>Death Date</td>
<td>Cause of Death</td>
<td>Vaccination Date</td>
<td>Patient ID</td>
<td>Administration Date &amp; Time</td>
</tr>
<tr>
<td>Source</td>
<td>Source</td>
<td>Admission Date</td>
<td>Administration Start &amp; End Date &amp; Time</td>
<td>Encounter ID</td>
</tr>
<tr>
<td>Confidence</td>
<td>Confidence</td>
<td>Vaccine Code &amp; Type</td>
<td>Encounter ID</td>
<td>National Drug Code (NDC)</td>
</tr>
<tr>
<td>Etc.</td>
<td>Etc.</td>
<td>Provider</td>
<td>Transfusion Administration ID</td>
<td>Route</td>
</tr>
</tbody>
</table>

## Clinical Data

<table>
<thead>
<tr>
<th>Lab Result</th>
<th>Vital Signs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient ID</td>
<td>Patient ID</td>
</tr>
<tr>
<td>Result &amp; Specimen Collection Dates</td>
<td>Measurement Date &amp; Time</td>
</tr>
<tr>
<td>Test Type, Immediacy &amp; Location</td>
<td>Height &amp; Weight</td>
</tr>
<tr>
<td>Logical Observation Identifiers Names and Codes (LOINC®)</td>
<td>Diastolic &amp; Systolic BP</td>
</tr>
<tr>
<td>Tobacco Use &amp; Type</td>
<td>Etc.</td>
</tr>
</tbody>
</table>

## Inpatient Data

<table>
<thead>
<tr>
<th>Inpatient Pharmacy</th>
<th>Inpatient Transfusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient ID</td>
<td>Patient ID</td>
</tr>
<tr>
<td>Administration Date &amp; Time</td>
<td>Administration Start &amp; End Date &amp; Time</td>
</tr>
<tr>
<td>Encounter ID</td>
<td>Encounter ID</td>
</tr>
<tr>
<td>National Drug Code (NDC)</td>
<td>Transfusion Administration ID</td>
</tr>
<tr>
<td>Route</td>
<td>Transfusion Product Code</td>
</tr>
<tr>
<td>Dose</td>
<td>Blood Type</td>
</tr>
<tr>
<td>Etc.</td>
<td>Etc.</td>
</tr>
</tbody>
</table>

## Mother-Infant Linkage Data

<table>
<thead>
<tr>
<th>Mother-Infant Linkage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mother ID</td>
</tr>
<tr>
<td>Mother Birth Date</td>
</tr>
<tr>
<td>Encounter ID &amp; Type</td>
</tr>
<tr>
<td>Admission &amp; Discharge Date</td>
</tr>
<tr>
<td>Child ID</td>
</tr>
<tr>
<td>Child Birth Date</td>
</tr>
<tr>
<td>Mother-Infant Match Method</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)
Differences among data models

- Comparison of 4 CDMs found OMOP the best for use with a longitudinal community registry

- Examples of differences
  - Derivative data and assumptions
    - Drug eras in OHDSI
  - Terminology binding
    - Commitment to standard terminologies (e.g., OHDSI)

- Ongoing harmonization (?)

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>
ICD-10 vs. ICD-10-CM

E72 Other disorders of amino-acid metabolism

Excl.: abnormal findings without manifest disease (R76.84)

- disorders of:
  - aromatic amino-acid metabolism (E70.1)
  - branched-chain amino-acid metabolism (E71.3)
  - fatty-acid metabolism (E71.3)
  - purine and pyrimidine metabolism (E79.3)
  - gout (M10.0)

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)
Unified Medical Language System

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

https://uts.nlm.nih.gov/
Integrating terminologies

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

Clinical repositories

Genetic knowledge bases

Biomedical literature

Addison's disease (363732003)

Other subdomains

SNOMED CT

OMIM

UMLS C0001403

NCBI Taxonomy

MESH

Model organisms

Anatomy

Genome annotations

FMA

GO

Biomedical literature

Addison Disease (D000224)
Current trends and future directions
ONC interoperability roadmap

Connecting Health and Care for the Nation
A Shared Nationwide 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.
ONC Interoperability Standards Advisory

◆ Introduction to the ISA
  ● Scope; Purpose; ISA Structure

◆ Sections
  ● I: Vocabulary/Code Set/Terminology Standards and Implementation Specifications
  ● II: Content/Structure Standards and Implementation Specifications
  ● III: Standards and Implementation Specifications for Services
  ● IV: Administrative Standards and Implementation Specifications

◆ Appendices
  ● I – Sources of Security Standards and Security Patterns
  ● II - Models and Profiles
  ● III - Educational and Informational Resources
  ● IV - State and Local Public Health Readiness for Interoperability

https://www.healthit.gov/isa/
Learning health system

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

http://www.fhir.org/
All of Us – Precision Medicine Initiative (NIH)

The future of health begins with you

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.

JOIN NOW