Research infrastructure of the Observational Health Data Sciences and Informatics (OHDSI) consortium: Institutional and researcher’s perspectives

MEDINFO 2017

Vojtech Huser
Christian Reich
Rae Woong Park
Speakers overview

- Vojtech Huser, MD, PhD
  - Staff scientist
  - National Institute of Health, U.S. National Library of Medicine

- Rae Wong Park, MD, PhD
  - Director of Korean Society of Medical Informatics
  - Director of Department of Biomedical Informatics, Ajou University School of Medicine

- Christian Reich, MD, PhD
  - VP Real World Insights, QuintilesIMS, USA
  - Principle Investigator OHDSI
Agenda

- Research network description
- Case studies
  - European implementation case study
  - Asian implementation case study
  - Data quality research study
- Questions
Agenda

- Research network description
- Case studies
  - European implementation case study
  - Asian implementation case study
  - Data quality research study
- Questions
History of OHDSI

- **OMOP (2008-2013)** [www.omop.org](http://www.omop.org)
  - OMOP = Observational Medical Outcomes Partnership
  - Research on methods for drug safety evaluation
    - Methods library developed; positive/negative drug outcome pairs
  - Common Data Model (then, was a byproduct)
  - Foundation for the NIH
    - Transition to Reagan Udall Foundation for the Food and Drug Administration

- **OHDSI (after 2013)** [www.ohdsi.org](http://www.ohdsi.org)
  - OHDSI = Observational Health Data Science and Informatics
  - Continues to use the name ‘OMOP CDM’
  - Community of researchers; public; non-pharma funded
Current Approach: “New, script based input data mapping for every study”

“What's the adherence to my drug in the data assets I can analyze?”

Analytical method: Adherence to Drug

Application to data

One SAS or R script for each study

- Not scalable
- Expensive
- Slow
- Prohibitive to non-expert routine use
Data Standardization Enables Systematic Research

Adherence

Mortality

Source of Business

Safety Signals

North America
Southeast Asia
China
Europe
UK
India
Japan
So Africa
Switzerland
Italy

Standardized data

OHDSI Tools

OMOP CDM
Standardized methods: ATLAS

- Design your study
  - What’s your target cohort?
  - What’s your compactor cohort?
  - What’s your outcome cohort?
  - What’s your time-at-risk?
  - What’s your model specification?
  - What’s your covariate adjustment strategy?

- Run
Analytics can be remote
Analytics can be behind firewall
Networks of networks

Coordinating Center

Another Network

Network

EMR Asset

EMR Asset

EMR Asset

EMR Asset

EMR Asset

EMR Asset

Claims Asset

Claims Asset

Claims Asset

Claims Asset

ISDN

University Medical Center

Inpatient Hospital

Outpatient Hospital
Past network studies

- **Clinical studies**
  - 2015
    - Treatment Pathway Study (diabetes, hypertension, depression)
  - 2016
    - Levetiracetam vs. phenytoin in epilepsy
    - Comparison of combination treatment in hypertension*
  - 2017
    - Sisyphus challenge (Alendronate vs. Raloxifene for osteoporosis)*
- **Other**
  - Anticoagulants, Prediction, Celecoxib vs. nsNSAIDs
- **Informatics studies**
  - 2015: Pediatric drug use epidemiology study
  - 2016: Achilles Heel Evaluation study
  - 2017: Data Quality
Focus of the panel

- Institutional perspective on OHDSI
  - Case studies 1 and 2

- Researcher’s perspective on OHDSI
  - Case study 3
Agenda

- Research network description
- Case studies
  - European implementation case study
  - Asian implementation case study
  - Researcher’s perspective: Data quality study
- Questions
Agenda

- Research network description
- Case studies
  - European implementation case study
  - Asian implementation case study
  - Researcher’s perspective: Data quality study
- Questions
Vojtech Huser
Agenda

- Research network description
- Case studies
  - European implementation case study
  - Asian implementation case study
  - Researcher’s perspective: Data quality study
- Questions
OHDSI: Researcher’s perspective

- **Strengths**
  - Analysis portability
    - Analysis written at one site can possibly be executed by other partners within the consortium
  - Common Data Model, OMOP Vocabularies
  - Tools + R packages
  - Community of researchers, past studies are open source

- **Weaknesses**
  - Must have recourses for data transformation to CDM
  - Expertise to install/use OHDSI tools and packages
Network results aggregation

- **Study conventions (R package) (STEP 1: Local Execution)**
  - `install_github("ohdsi/StudyProtocolSandbox/DataQuality")`
  - `executeDQ(connectionDetails, cdmDatabaseSchema, resultsDatabaseSchema, workFolder='c:/mystudy')`

- **Package results (STEP 2)**
  - .zip file which a site researcher inspects closely
  - `install_github("ohdsi/OhdsiSharing")`
  - `packageResults(…, workFolder, dbName)`
  - `submitResults(…, studyBucketName, studyKey, studySecret)`

- **STEP 3: Aggregated data analysis**

- **Full example**
  - [https://github.com/OHDSI/StudyProtocolSandbox/tree/master/DataQuality#2participate-on-dataquality-study](https://github.com/OHDSI/StudyProtocolSandbox/tree/master/DataQuality#2participate-on-dataquality-study)
Data logistics (example)
OMOP Vocabularies

- Common framework
  - CONCEPT, CONCEPT_RELATIONSHIP, CONCEPT_ANCESTOR
- Benefit: pre-build infrastructure [+ mapping]
- Browser
  - http://www.ohdsi.org/web/atlas/#/concept/21600381

- Example of a researcher benefit
  - ICD10CM -> SNOMED CT (after Oct 1st, 2015)
  - ICD9CM -> SNOMED CT
# Acute renal failure syndrome

<table>
<thead>
<tr>
<th>Id</th>
<th>Code</th>
<th>Name</th>
<th>Class</th>
<th>RC</th>
<th>DRC</th>
<th>Distance</th>
<th>Domain</th>
<th>Vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td>44833292</td>
<td>283.11</td>
<td>Hemolytic-uremic syndrome</td>
<td>5-digit billing code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD9CM</td>
<td>ICD9CM</td>
</tr>
<tr>
<td>44825529</td>
<td>572.4</td>
<td>Hepatorenal syndrome</td>
<td>4-digit billing code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD9CM</td>
<td>ICD9CM</td>
</tr>
<tr>
<td>44833658</td>
<td>S84</td>
<td>Acute kidney failure</td>
<td>3-digit nonbill code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD9CM</td>
<td>ICD9CM</td>
</tr>
<tr>
<td>44837189</td>
<td>S84.6</td>
<td>Acute kidney failure with lesion of renal cortical necrosis</td>
<td>4-digit billing code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD9CM</td>
<td>ICD9CM</td>
</tr>
<tr>
<td>44820369</td>
<td>S84.8</td>
<td>Acute kidney failure with other specified pathological lesion in kidney</td>
<td>4-digit billing code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD9CM</td>
<td>ICD9CM</td>
</tr>
<tr>
<td>44826731</td>
<td>S84.9</td>
<td>Acute kidney failure, unspecified</td>
<td>4-digit billing code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD9CM</td>
<td>ICD9CM</td>
</tr>
<tr>
<td>44823297</td>
<td>669.3</td>
<td>Acute kidney failure following labor and delivery</td>
<td>4-digit nonbill code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD9CM</td>
<td>ICD9CM</td>
</tr>
<tr>
<td>44827985</td>
<td>669.30</td>
<td>Acute kidney failure following labor and delivery, unspecified as to episode of care or not applicable</td>
<td>4-digit billing code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD9CM</td>
<td>ICD9CM</td>
</tr>
<tr>
<td>44837293</td>
<td>669.32</td>
<td>Acute kidney failure following labor and delivery, delivered, with mention of postpartum complication</td>
<td>5-digit billing code</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>Condition ICD9CM</td>
<td>ICD9CM</td>
</tr>
<tr>
<td>44824443</td>
<td>669.34</td>
<td>Acute kidney failure following labor and delivery, postpartum condition or complication</td>
<td>5-digit billing code</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>Condition ICD9CM</td>
<td>ICD9CM</td>
</tr>
<tr>
<td>35206734</td>
<td>D59.3</td>
<td>Hemolytic-uremic syndrome</td>
<td>4-char billing code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD10CM</td>
<td>ICD10CM</td>
</tr>
<tr>
<td>35208366</td>
<td>K76.7</td>
<td>Hepatorenal syndrome</td>
<td>4-char billing code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD10CM</td>
<td>ICD10CM</td>
</tr>
<tr>
<td>45552606</td>
<td>K91.82</td>
<td>Postprocedural hepatorenal syndrome</td>
<td>5-char billing code</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>Condition ICD10CM</td>
<td>ICD10CM</td>
</tr>
<tr>
<td>1571485</td>
<td>N17</td>
<td>Acute kidney failure</td>
<td>3-char nonbill code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD10CM</td>
<td>ICD10CM</td>
</tr>
<tr>
<td>35209270</td>
<td>N17.1</td>
<td>Acute kidney failure with acute cortical necrosis</td>
<td>4-char billing code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD10CM</td>
<td>ICD10CM</td>
</tr>
<tr>
<td>35209272</td>
<td>N17.8</td>
<td>Other acute kidney failure</td>
<td>4-char billing code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD10CM</td>
<td>ICD10CM</td>
</tr>
<tr>
<td>35209273</td>
<td>N17.9</td>
<td>Acute kidney failure, unspecified</td>
<td>4-char billing code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD10CM</td>
<td>ICD10CM</td>
</tr>
<tr>
<td>35210387</td>
<td>C90.04</td>
<td>Postpartum acute kidney failure</td>
<td>4-char billing code</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>Condition ICD10CM</td>
<td>ICD10CM</td>
</tr>
</tbody>
</table>
Learning from past studies

- Evolution over time
  - increasing sophistication (SQL, SQL+R packages, portable phenotypes)
- Anti-epilepsy drug analysis (levetiracetam) (“second generation”)

Risk of angioedema associated with levetiracetam compared with phenytoin: Findings of the observational health data sciences and informatics research network.

Duke JD1,2, Ryan PB1,3,4, Suchard MA1,5, Hripcsak G1,4, Jin P1,4, Reich C1,6, Schwalm MS1,6, Khoma Y1,7,8, Wu Y1,9, Xu H1,9, Shah N1,10, Banda JM1,10, J Schuemie M1,3.

- https://github.com/OHDSI/StudyProtocols
- https://github.com/OHDSI/StudyProtocolSandbox/
Study package (in R language)

OHDSI Keppra and the Risk of Angioedema study
Value Set definition ("second generation" study example)

```sql
select codeset_id, concept_id
INTO #Codesets
FROM
(
    SELECT 0 as codeset_id, c.concept_id FROM (select distinct l.concept_id FROM
    (select DISTINCT concept_id from @cdm_database_schema.CONCEPT where concept_id in (711584) and invalid_reason is null
    UNION
    SELECT 1 as codeset_id, c.concept_id FROM (select distinct l.concept_id FROM
    (select DISTINCT concept_id from @cdm_database_schema.CONCEPT where concept_id in (40294488) and invalid_reason is null
    UNION
    SELECT c.concept_id FROM @cdm_database_schema.CONCEPT c
    join @cdm_database_schema.CONCEPT_Ancestor ca on c.concept_id = ca.descendant_concept_id
    and ca.ancestor_concept_id in (711584)
    and c.invalid_reason is null
    ) l
    ) c
    UNION
    SELECT 1 as codeset_id, c.concept_id FROM (select distinct l.concept_id FROM
    (select DISTINCT concept_id from @cdm_database_schema.CONCEPT where concept_id in (40294488) and invalid_reason is null
    UNION
    SELECT c.concept_id FROM @cdm_database_schema.CONCEPT c
    join @cdm_database_schema.CONCEPT_Ancestor ca on c.concept_id = ca.descendant_concept_id
    ) l
    ) c
```

“Third generation” study

Comparative effectiveness of alendronate and raloxifene in reducing the risk of hip fracture

Objective: To compare the effectiveness in reducing the risk of hip fracture between alendronate and raloxifene.

Rationale: Osteoporosis is characterized by decreased bone mass and deterioration of bone tissue, resulting in reduced bone strength and increased fracture risk. Approved therapies for osteoporosis include bisphosphonates, calcitonin, raloxifene and teriparatide. Among these drugs, alendronate and raloxifene are the most popular osteoporosis medication and a burden of prescription are performed annually.
### Third generation study

#### Cohorts To Create

<table>
<thead>
<tr>
<th>cohortId</th>
<th>atlasId</th>
<th>name</th>
</tr>
</thead>
<tbody>
<tr>
<td>99321</td>
<td>99321</td>
<td>Alendronate</td>
</tr>
<tr>
<td>99322</td>
<td>99322</td>
<td>Raloxifene</td>
</tr>
<tr>
<td>99323</td>
<td>99323</td>
<td>HipFracture</td>
</tr>
<tr>
<td>100791</td>
<td>100791</td>
<td>VertebralFracture</td>
</tr>
<tr>
<td>100792</td>
<td>100792</td>
<td>NonHipNonVertebralFracture</td>
</tr>
<tr>
<td>100793</td>
<td>100793</td>
<td>OsteonecrosisOfJaw</td>
</tr>
<tr>
<td>100794</td>
<td>100794</td>
<td>EsophagealCancer</td>
</tr>
<tr>
<td>100795</td>
<td>100795</td>
<td>AtypicalFemoralFracture</td>
</tr>
</tbody>
</table>
Web-based phenotype definition

Cohort definition: A cohort is defined as the set of persons satisfying one or more inclusion criteria for a duration of time. One may qualify for one cohort multiple times during non-overlapping time intervals. Cohorts are constructed in ATLAS by specifying cohort entry criteria and cohort exit criteria. Cohort entry criteria involve selecting one or more initial events, which determine the start date and optionally specifying additional inclusion criteria which filter to the qualifying events. Cohort exit criteria are applied to each record to determine the end date when the person’s episode no longer qualifies for the cohort.

Initial event cohort: Events are recorded time-stamped observations for the persons, such as drug exposures, conditions, procedures, measurements and visits. All events have a start date and end date, though some events may have a start date and end date with a value (such as procedures or measurements). The event index date is set to be equal to the event start date.

People having any of the following: Add Initial Event...

- a drug exposure of: alendronate
- for the first time in the person’s history
- occurrence start is: Before 2012-02-01

with continuous observation of at least 365 days before and 90 days after event index date

Limit initial events to: earliest event per person.
Data Quality Study

- 12+ datasets (from 7 sites)
  - dataset metadata (least aggressive data view)
    - re-using dataset characterization pre-computations (from Achilles OHDSI tool;
    - Number of distinct procedure concepts per person
  - Ethical review; US: IRB (=institutional review board)
    - OHDSI central IRB

- Empirical comparison
  - DQ rules vs. empirical thresholds
    - % of patients with at least one visit

<table>
<thead>
<tr>
<th>ach_2003:Percentage (1+ visit)</th>
<th>median</th>
<th>percentile10</th>
<th>min</th>
<th>max</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>89.96</td>
<td>62.74</td>
<td>37.82</td>
<td>100</td>
</tr>
</tbody>
</table>
Unmapped data (example #1)

- Count of local codes not mapped to a standard
Unmapped data (example #2)

- Percentage of unmapped data by domain
Questions
Agenda

- Research network description
- Case studies
  - European implementation case study
  - Asian implementation case study
  - Researcher’s perspective: Data quality study
- Questions