A success story

The Adoption of the OMOP-CDM in South Korea

Rae Woong Park MD, PhD
Ajou University School of Medicine South Korea
Disclosures

Founder and stockholder of EvidNet

- OMOP CDM-based distributed research network platform startup company in Korea
- One-step CDM conversion service: legacy EMR data to OMOP CDM
- Vocabulary mapping service: mapping of local vocabulary to OMOP standard vocabulary
- Maintenance service: routine maintenance of CDM DB (ex) CDM version/vocabulary updates, integration of new CDM extensions, etc.
- Distributed research network platform for participating hospitals (in development)
Contents

01 About South Korea
02 Korean OHDSI Network
03 Why is DRN and CDM popular in Korea?
04 Lessons learned from potential Data Owner
05 Future plans in Korean OHDSI
About South Korea
Selected Indicators for Korea

GDP (nominal) 2017 estimate
- Total: $1,498 trillion (11th)
- Per capita: $29,114 (27th)

Life expectancy
- Men: 77 years
- Women: 84 years

Population
- 50.4 million
- Growth rate: 0.4% ↓

Source: OECD DATA
Hospitals in Korea

- About 43 University or Tertiary Teaching Hospitals
- 280 general hospitals (>100 beds)
- 2,800 hospitals (>30 beds)
Status of Korean OHDSI
Status of Korean OHDSI Network

Data Network of 39 Hospitals, 54M Patients

Seoul

Incheon / Gyeonggi

Chungcheong

Gangwon

Jeolla

Gyeongsang
Commitment by 39 Hospital Heads to Share IRB Review Results and Conduct Joint Research

Data Network

39 Hospitals, 54M Patients
CDM Conversion History

2014
Ajou University Hospital
CDM Complete (First in Korea)

2015
Gachon University Gil Hospital
CDM Complete

2016.10
Korea National Health Insurance Service (NHIS)
CDM Conversion of 1M Patients

2017.09
Korea National Health Insurance Service (NHIS)
CDM Conversion of 1M Patients

2017.10
Gangwon State University Hospital
CDM Complete

2017.11
Korea Health Insurance Review and Assessment Service (HIRA)
CDM Conversion of 1M CV Patients

2018.01
Samsung Medical Center, Wonkwang University Hospital, Wonju Severance Christian Hospital, Chonbuk National University Hospital
CDM Complete (4 Hospitals)
Why is DRN and CDM popular in Korea?
Invited Talks

226 invited talks during past 42 months since July 2014

- 2014: 13 times
- 2015: 42 times
- 2016: 74 times
- 2017: 92 times
Number of invited presentations

79% Increase

- 2014: 13
- 2015: 42
- 2016: 74
- 2017: 92
Lesson learned from potential Data Owner

- Quick-prototyping
- Live demonstration
- Success story
- Focusing on clinicians
- Data governance leadership meeting
Strategy for Standard Vocabulary Mapping

- Sort codes by frequency of usage
- Number of codes required to cover

95% of transaction data → 99%
20-30% of codes → 60-70% of codes
Achilles: data characterization

Distributed research network

Openness
http://ami.ajou.ac.kr:8080
Examples of clinical researches using big data

Characterizing treatment pathways at scale using the OHDSI network

George Hiripcasas, Patrick B. Ryan, Jon D. Duke, Nigara H. Shah, Rae Wooong Park, Vtgiris Hoser, Marc A. Suchan, Martin J. Schuemie, Frank J. DeFalco, Adler Perotte, Juan M. Banda, Christian G. Reich, Lisa M. Schilling, Michael E. Matheny, Daniella Mestek, Nicole Pratt, and David Madigan

*Department of Biomedical Informatics, Columbia University Medical Center, New York, NY 10032; 2Medical Informatics Service, NewYork-Presbyterian Hospital, New York, NY 10032; 3Biomedical Data Science and Informatics, New York, NY 10032; 4Nursing Informatics, New York, NY 10032; 5Janssen Research and Development, Raritan, NJ 07786; 6Center for Biomedical Informatics, NewYork-Presbyterian Hospital, New York, NY 10032; 7Department of Biomedical Informatics, Aga Khan University School of Medicine, Karachi, Pakistan; 8Department of Informatics, University of California, Los Angeles, CA 90095; 9Institute of Informatics, University of California, Los Angeles, CA 90095; 10Department of Biomedical Informatics, University of California, Los Angeles, CA 90095; 11Department of Biomedical Informatics, University of California, Los Angeles, CA 90095; 12Quiring Library, University of California, Los Angeles, CA 90095; 13Health Sciences Library, University of California, Los Angeles, CA 90095; 14Biostatistics, University of Southern California, Los Angeles, CA 90007; 15Department of Pediatrics, University of Southern California, Los Angeles, CA 90007; 16Division of Health Sciences, University of Southern California, Los Angeles, CA 90007; and 17Department of Biostatistics, Columbia University, New York, NY 10032

Observed research promises to complement experimental research by providing large, diverse populations that would be unfeasible for an experiment. Observational research can test its own clinical hypotheses, and observational studies also can contribute to the design of experiments and inform the generalizability of experimental research. Understanding the diversity of populations and the variance in care is one component. In this study, the Observational Health Data Sciences and Informatics (OHDSI) collaboration created an international data network with 11 data sources from four countries, including electronic health record and administrative datasets. Without sufficiently broad datasets available in the first stage, randomized trials are designed without explicit knowledge of actual disease status and treatment practice. Literature reviews are restricted to the population choices of previous investigations, and pilot studies usually are limited in scope. By exploiting the ClinicalTrials.gov national trial registry (9) and electronic health record, researchers already have demonstrated the discrepancy between targeted populations and populations available for study (10), raising the concern that designs may not be optimal. Designs cannot be based simply on current treatment recom
Lead/participation in CDM-based international joint researches (≥14)

- Season-Wide Association Study
  Boland, Mary Regina, Columbia University
  3 Countries, 6 DBs, 10M Patients

- Characterizing treatment pathways at scale using the OHDSI network
  George Hripcsak, Columbia University
  5 Countries, 11 DBs, 250M Patients

- Multisite Evaluation of a Data Quality Tool for Patient-Level Clinical Data Sets
  Vojtech Huser, NIH
  2 Countries, 24 DBs

- Applying a Common Data Model to Asian Databases for Multinational Phase-IV Pharmacovigilance Studies: Opportunities and Challenges
  Ai-Wen Chiou, Chinese University of Hong Kong, University of Taiwan
  2 Countries, 24 DBs

- Learning Effective Treatment Pathways
  Ngan Shin, Stanford University
  3 Countries, 7 DBs, 100M Patients

- Comparative safety and effectiveness of alemtuzumab versus raloxifene in women with osteoporosis
  Marc Suchard, UCLA
  2 Countries, 9 DBs, 280M Patients

- Hypertension Research
  Led by Ajou University

- 2017 OHDSI Symposium
  Best Practice Award
**Focusing on clinicians**

**Clinicians, rather than informaticians**

Major decision makers in a hospital are usually clinicians.

**Young clinicians assistant – associate professors**

Highest interest on OMOP CDM, because they do not have enough fund, resources and data for their research.

**Inter-disciplinary clinical meeting**

Then homogenous clinical meeting at initial stage.
Ajou University has been coordinating OHDSI in Korea since March 2017 and is actively initiating democratic healthcare data governance discussions with CIOs from hospitals of various sizes.

- Ajou University Hospital 2017.03.07
- Samsung Medical Center 2017.05.31
- Gangdong Sacred Heart Hospital (Hallym University) 2017.07.27
- Ewha Womans University Mokdong Hospital 2017.09.21
- Korea University Anam Hospital 2017.11.09
- Korea University Guro Hospital 2018.01.25
Activity in Korea

Engineer Meeting

- Biweekly, TC (current 19th)
- EHR experts from participating hospitals
- Discuss all the technical issues during CDM conversion

Open Forum

- Monthly, 1-3 hour lecture
- Agenda
  - Introduction to OHDSI and CDM
  - OMOP CDM Structure
  - OMOP CDM Vocabulary/ vocabulary mapping
  - Tools for OMOP CDM
  - ETL process
  - Research Experience using OHDSI Network
Government’s interest in CDM
No. of Presentations and Data Partner Join

- Presentations
- Join

18 Months

Data Partners

Gacheon

NHIS

Year

2014 2015 2016 2017

Presentations

200
150
100
50
0
NHIS, Development of a CDM-based Drug Safety Surveillance System

Three year project 2016-2018

1st year | Feasibility
9-year of 1M pt data into CDM

2nd year | Validation
validation of usefulness of the CDM

3rd year | Full conversion
12-years of 51 M patients

OMOP CDM conversion

RAW data
SAS file?

ETL Layer
Voca mapper
ETL
Data Conversion

National Health Insurance Service: Governmental National Health Insurer

Application of pharmacovigilance tools

PV tools
ROR, PRR, LGPS
CLEAR CERT

OHDSI tools
ATLAS
ACHILLES
### Project Results

#### Designing an Efficient System to Support CDM Conversion of the Entire Korean Population

**Simulation of required system specifications**

*based on Oracle Exadata*

- Uncompressed CDM data of 1M patients (2%) = 527GB
- Uncompressed CDM data of 50M patients (100%) = 25.8TB (estimate)
- Total 103TB (x4) required considering index and temp domain

<table>
<thead>
<tr>
<th>Table Name</th>
<th>Column Size (Byte)</th>
<th>Data Count</th>
<th>Total Size (MB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Patient Information</td>
<td>261</td>
<td>999,812</td>
<td>248.9</td>
</tr>
<tr>
<td>Visit Information</td>
<td>152</td>
<td>224,085,191</td>
<td>32,483.10</td>
</tr>
<tr>
<td>Location Information</td>
<td>400</td>
<td>317</td>
<td>0.1</td>
</tr>
<tr>
<td>Site Information</td>
<td>406</td>
<td>188,814</td>
<td>73.1</td>
</tr>
<tr>
<td>Drug Exposure Information</td>
<td>344</td>
<td>406,670,758</td>
<td>133,414.00</td>
</tr>
<tr>
<td>Continuous Drug Exposure Period</td>
<td>10</td>
<td>154,502,958</td>
<td>1,473.50</td>
</tr>
<tr>
<td>Diagnosis Information</td>
<td>181</td>
<td>355,902,525</td>
<td>61,434.10</td>
</tr>
<tr>
<td>Continuous Diagnosis Period</td>
<td>10</td>
<td>125,949,247</td>
<td>1,201.10</td>
</tr>
<tr>
<td>Observation Information</td>
<td>521</td>
<td>186,289,956</td>
<td>92,560.80</td>
</tr>
<tr>
<td>Death Information</td>
<td>506</td>
<td>45,767</td>
<td>22.1</td>
</tr>
<tr>
<td>Observation Period Information</td>
<td>9</td>
<td>999,813</td>
<td>8.6</td>
</tr>
<tr>
<td>Insurance Coverage Period</td>
<td>158</td>
<td>999,812</td>
<td>150.7</td>
</tr>
<tr>
<td>Procedure Information</td>
<td>160</td>
<td>961,582,127</td>
<td>146,725.80</td>
</tr>
<tr>
<td>Measurement Information</td>
<td>283</td>
<td>260,294,988</td>
<td>70,251.00</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3,401</strong></td>
<td><strong>2,678,512,085</strong></td>
<td><strong>540,046.80</strong></td>
</tr>
</tbody>
</table>
Project Results
Drug Use Analysis System Based on NHIS CDM Data

Roadmap

**CDM Infrastructure**

- **Level 1 (2016)**
  - Pilot
  - Development of claim data CDM conversion technology
  - 1M sample data conversion

- **Level 2 (2017)**
  - Validation
  - Code mapping update and data quality management
  - Calculation of required system specifications

- **Level 3 (2018~)**
  - Scale-up
  - CDM conversion of entire Korean population (50M)
  - CDM table extension
  - System environment setup

**CDM Analysis System**

- **Level 1 (2016)**
  - Pilot
  - Causal inference of drug-induced AEs

- **Level 2 (2017)**
  - Validation
  - Validation of analysis system and tool feasibility
  - Improvement of causal inference analysis

- **Level 3 (2018~)**
  - Launch
  - Automation of large-scale analysis
  - Trend analysis of data usage
  - Monitoring of disease indicators
  - Cost effectiveness analysis

**Level 4 (2018~)**

- Expansion of CDM Data Application
  - Promotion and training
  - Big data analysis environment setup
  - Development and application of various research methodologies
  - Development of tools such as drug-induced AE prediction, visualization of analysis results, etc.
  - Continuous CDM maintenance
Project Results

Drug Use Analysis System Based on NHIS CDM Data

- **External Research Assistance**
  - Research on therapeutic effect vs. AE
  - Usage analysis of drugs, etc.
  - Personalized disease predictions
  - Reference data for policy research
  - Reference data for new drug development
  - Research on international disease and treatment trends
  - Development of research methodologies and analytic tools

- **Information for Public**
  - Provide information about drugs, disease, etc.
  - Provide information on potential high risk groups
  - Provide information on reimbursed drugs

- **Application within NHIS**
  - Provide resources related to drug pricing
  - Generate background data to improve reimbursement
  - Big data analysis training

- **CDM Applications**
  - Information for Public
  - External Research Assistance
Bio & Medicine
Korea to build database for medical big data

PUBLISHED February 09, 2018 - 17:13   UPDATED February 09, 2018 - 17:13

[THE INVESTOR] The South Korean government will consolidate medical data from hospitals across the country and establish a comprehensive database, as part of its goal to foster the country’s lucrative pharmaceutical and healthcare industries, officials said on Feb. 9.

Its extended vision is to create some 35,000 new jobs in the growing bio and healthcare sector and to pull up the nation’s global market share from the current 1.8 percent to 4 percent by 2022.

The Ministry of Trade, Industry and Energy on Feb. 9 held a conference with local bio & healthcare players at Wonju Medical Industry Techno Valley and announced the government’s short- and mid-term development strategies for the industry.

The event was attended by officials from Daewoong Pharm, Mediana Samsung Seoul Hospital, Korea Biotechnology Industry Organization and other local medical organizations.

As a pilot step, the ministry is to select six major hospitals within the first half of the year and collect genetic and biometric data of some 10 million people to create a bio database on bio information, according to officials.
Distributed Bio Big Data Model

Bio Big Data Center

- Analysis SW
- Analysis Results
- Analysis request
- Analysis Results
- Analysis SW request
- Analysis SW supply

Hospital

Company

IT Company
Ministry of Industry - Ministry of Health and Welfare 'Bio-Big Data'

Enter the article 2017-04-20 10:16

[Ellection Day Sejong = Park Eally]

Ministry of Health and Welfare launches big medical data promotion group in March ... The Ministry of Industry is also promoting TF

In the era of the Fourth Industrial Revolution, in order to preoccupy the bio-data field, there is a fierce nervous battle between government departments.

According to the Ministry of Health and Welfare, the Ministry of Health and Welfare, which is in charge of the hospitals and the pharmaceutical industry, started a pilot project last year to build and utilize healthcare data. Also, last month, the 'Big Health Data Medical Promotion Team' was launched in order to create a strategy to utilize big medical data. The Ministry of Industry, Trade and Industry and the Future Creation Science Department also participated in the promotion team, and the head of the Health and Medical Policy Division of the Kim Gang - The Ministry of Health and Welfare is planning to establish a platform to link big data of public institutions such as National Health Insurance Corporation and Health Insurance Review and Evaluation Center.
Collaboration and Competition in between Government CDM

Ministry of Science and ICT

Cloud HIS with CDM

Ministry of Food and Drug Safety

DRN, OMOP-CDM
Claim + Health Exam
DRN, K-CDM (?)
EMR

h·well
National Health Insurance Corporation

Ministry of Trade, Industry and Energy

DRN, OMOP-CDM
EMR + Omics + life-log

DRN, OMOP-CDM
EMR + Claim
HIRA, KCDC,
National Cancer Center

Ministry of Health and Welfare
Characteristics of Korean OHDSI

**Data partners**
- Major tertiary teaching hospitals
  - Detailed time stamp
  - Test results
  - Outcome data

**National Health Insurance Data**
- Compulsory health insurance
- Claim data
  + Socioeconomic data
  + Regular Health exam data (includes lab tests)
- 10-year of observation period
- Covers all the citizens (50M)
CDM-Based Distributed Data Network

Development of an OMOP CDM-Based Distributed, Integrative Bio Health Data Network
Request for Proposal Overview

Final Goal
- To develop key technologies required to setup and operate an OMOP CDM-based data network
- To promote industrial application of hospital data such as clinical data, genomics data, biosignal data, etc.

- Development and validation of CDM extensions for genomics and biosignal data
- Conversion of EMR data to OMOP CDM
- Quality management of CDM data
- Development of analytical tools for CDM data
- Development of information security technologies

Period  Within 36 months (9 months in 2018)
Grant  Total $4M ($0.8M in 2018)
Host Organization  No restrictions
* Participation of hospitals and small/midsize businesses required

- Participating hospitals must agree to share IRB review results and conduct joint research (commitment from hospital heads required)
- Participating businesses must present respective business models based on CDM

Development Goals

<table>
<thead>
<tr>
<th>Evaluation Criteria</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Development of CDM extensions (genomics, lifelog, etc.)</td>
<td>≥ 2</td>
</tr>
<tr>
<td>2  Participating hospitals</td>
<td>≥ 6</td>
</tr>
<tr>
<td>3  Patients converted to CDM</td>
<td>≥ 1M</td>
</tr>
<tr>
<td>4  Joint researches</td>
<td>≥ 5</td>
</tr>
<tr>
<td>5  Business models based on CDM data</td>
<td>≥ 3</td>
</tr>
<tr>
<td>6  Development of data quality assurance tools for CDM</td>
<td>≥ 3</td>
</tr>
</tbody>
</table>
Proposal Outline

Coordinating Platform

Hospitals (39, 34k beds)

Patient Data (54M)

Extension Model (7)

Research Institutes (10)

Companies (7)

BM (7)
Collaborating or Participating Organizations

CLOVER™
Global Collaborating Organizations
Research Institutes/Companies in Demand

AI-CTS
Smart Clinical Research Navigator
PHI Protection

CLOVER™ Node (Participating Hospital)

01 Verification of Sensitive Data
Detection and de-identification of PHI in compliance to HIPAA

02 Verification of Code Security
Detection of malware prior to query code import

03 Detection of PHI
Detection of PHI based on k-anonymity, l-diversity and t-closeness prior to query result export
Personalized Services

Hospital
- EMR
- OMOP / CDM

Company
- Life-log
- Genomic ...

Data Interface
- Platform
- Deep Learning Engine
- Analysis Engine

- Genomics CDM Extension Based
  BM Discovery
- Genomic Biomarker and Resistance Mutation Based
  Personalized Oncology
- Chronic Disease High-Risk Group
  Care Service

Services based on genomic + clinical data
- Patent application of platform technology and service: > 2
- Commercialization: > 2

Chronic Disease Care Service
Perinatal Mother Care Service

PhR Platform
- OPEN API

Bio CDM Platform
- Discovery of Cancer-specific Mutations
- CDS Gallery
- Gene mutation-protein structure-drug correlation visualization
- Treatment recommendation based on mutations

Dietary and prescription service for chronic disease patients after discharge
Pharma R&D Services

Hospital
- EMR
- OMOP / CDM

Company
- Life-log
- Genomic
- DATA BASE

Data Interface
- Platform
- Deep Learning Engine
- Analysis Engine

Clinical Trial Eligible
Patient Identification
- Deep EMR System
- Phenotyping of patients/disease
- Predictive modeling for practice
- Patient identification for clinical trials

Fixed Dose Combination (FDC) Recommendation & Drug Repositioning
- CDM. Pharma System
  - Big data analysis, machine learning
  - Drug-induced AE analysis
  - Co-prescription pattern and patient outcome analysis
  - Patient outcome analysis based on prescription schedule
  - Discovery of unmet needs

Results
- Incrementally modified drugs (IMDs) with improved safety profile
- New FDCs
- Discovery of niche markets
- Protocol optimization and feasibility assessment
Summary

01 About South Korea

02 Korea OHDSI Network

03 Why is DRN and CDM popular in Korea?

04 Lesson learned from potential Data Owner

- Quick-prototyping
- Live demonstration
- Success story
- Focusing on clinicians

05 Future plans in Korean OHDSI
Contacts  Rae Woong Park

- E-mail : veritas@ajou.ac.kr
- Tel : +82-31-219-4471