



# OHDSI/OMOP Introduction

Mui Van Zandt

IQVIA

VP/Global Head, Data Strategy, Access, and Enablement



# Instructors



Mui Van Zandt  
IQVIA



Mukkesh Kumar  
A\*STAR



Erwin Tantoso  
A\*STAR



Evelyn Goh  
National University of Singapore



Gyeol Song  
IQVIA



Seng Chan You  
Yonsei University



Thanh-Phuc Phan  
University Medical Center HCMC  
University of South Australia



Nicole Pratt  
University of South Australia

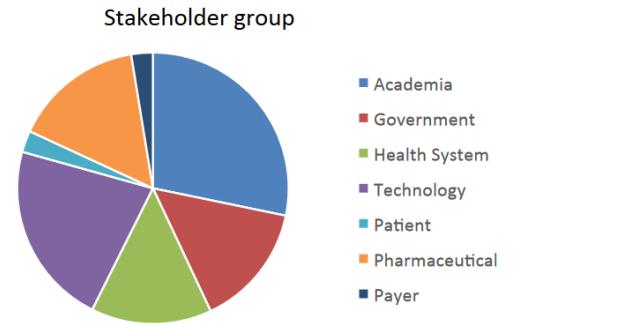


# Agenda

Session Name	Time	Instructor(s)
OHDSI/OMOP Intro	09:00 – 09:20	Mui Van Zandt
OMOP CDM and Vocabulary	09:20 – 10:00	Mukkesh Kumar & Erwin Tantoso
OMOP Conversion Process	10:00 – 10:30	Evelyn Goh
Energy Break	10:30 – 10:40	
ETL Exercises	10:40 – 12:00	Gyeol Song
Lunch	12:00 – 13:30	
OHDSI Analyses: Building Cohorts	13:30 – 14:50	Seng Chan You
CohortDiagnostics & Population-Level Estimation	14:50 – 15:30	Thanh-Phuc Phan
Interpreting Results	15:50 – 16:30	Nicole Pratt

## What OHDSI is:

- ✓ Open Source
- ✓ Community
- ✓ Data



## Why Choose OHDSI/OMOP:

- ✓ **Fast, reliable** studies across a series of datasets and data types
- ✓ **Reduced cost of ownership** including understanding coding schemes, writing statistical programs across databases or developing software
- ✓ **Expanded data access** via the OHDSI network and remote multi-center database studies



### OHDSI Collaborators:

- 4,751 collaborators
- >1,100 organizations
- 88 countries from 6 continents

### OHDSI Network:

- 544 data sources
- 974M unique patient records

<https://ohdsi.org/>



# OHDSI's Mission

To improve health by **empowering** a community to **collaboratively** generate the evidence that promotes better health decisions and better care.

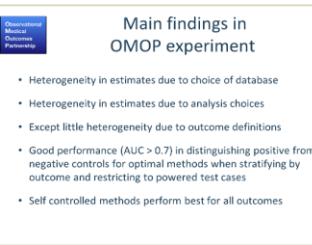


# History of OMOP/OHDSI

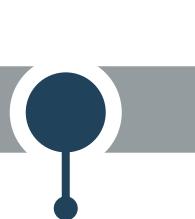
Global Acceptance



## End of OMOP Experiment

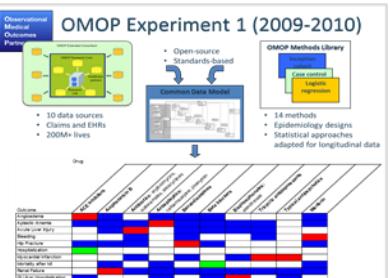


2013



2009

## OMOP Experiment #1



## First OHDSI Symposium/ Network Study Published

## First Hackathon at Columbia University

2015



## Launch of OHDSI



## Korea Chapter



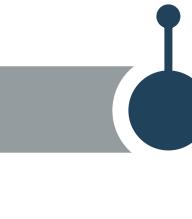
## China Chapter



## European Chapter

## FDA Adoption (FDA BEST Launch)

2017



2016

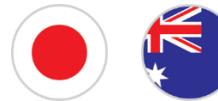
## EHDEN Initiation (Europe)

## FEEDER-NET Initiation (Korea)

## First European Symposium

## EMA Adoption

## Australia, Japan Chapters



2019



## OHDSI COVID-19 Study-a-Thon

## Singapore Chapter



## India Chapter



2022

## OMOP in Thailand

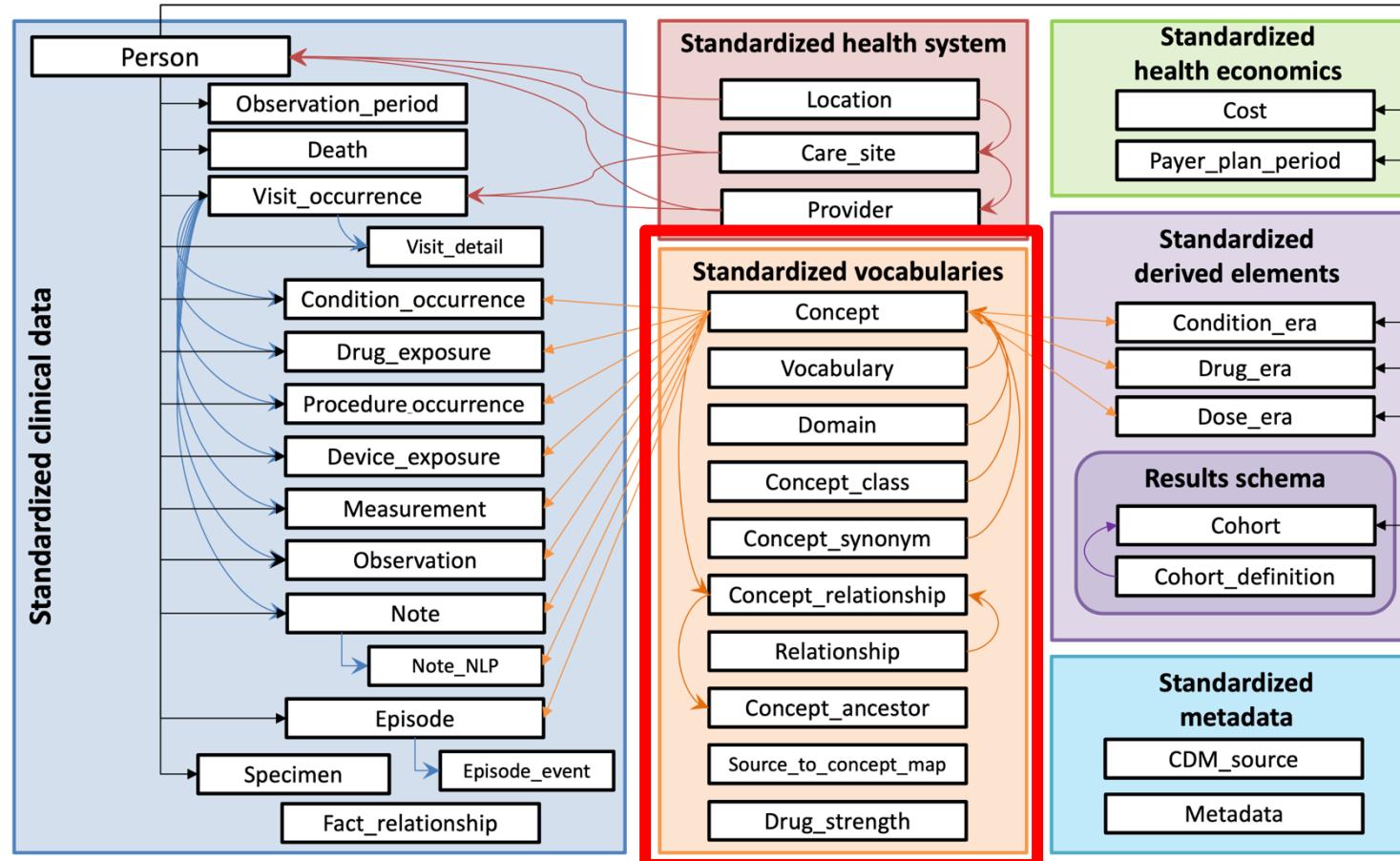
**Thailand (1)**  
Siriraj Hospital EHR

2023



# OMOP Common Data Model (CDM)

*Ontologies are critical when designing data models*



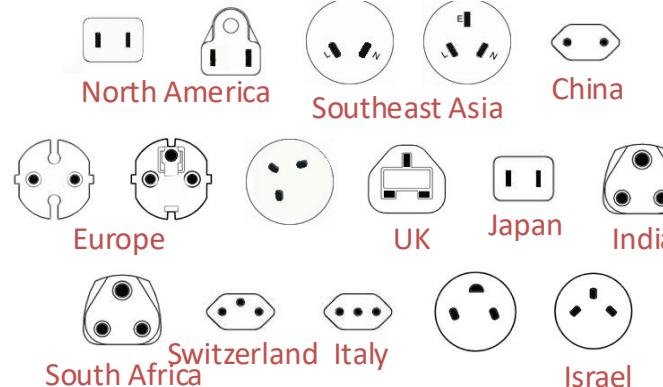
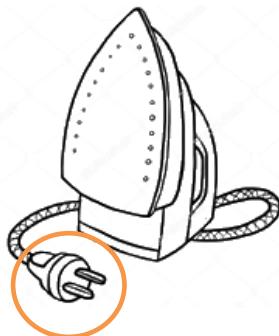
As of August 2025

- **11,804,307 concepts**
  - 3,784,263 standard concepts
  - 971,914 classification concepts
- **145 vocabularies**
- **43 domains**
- **87,948,636 concept relationships**
- **101,696,159 ancestral relationships**
- **6,028,711 concept synonyms**

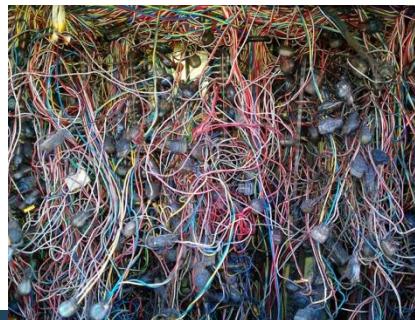
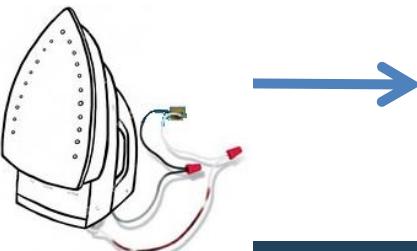


# Data Standardization to OMOP Enables Systematic Research

Analytical method:  
Adherence to Drug

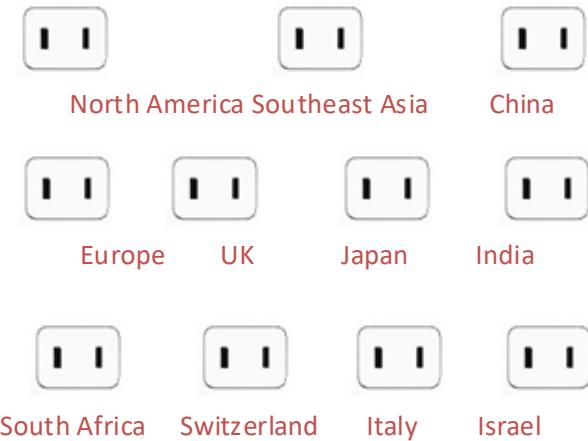


One SAS or R script  
for each study

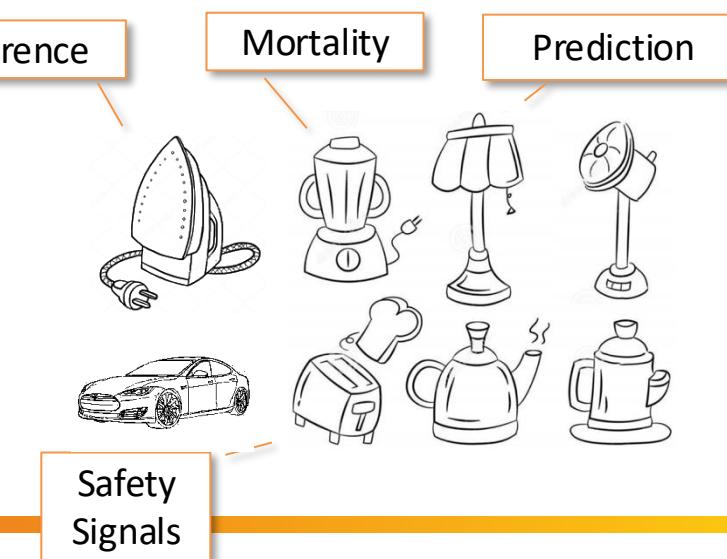


- Reliant on partner capabilities
- Not scalable
- Not transparent
- Expensive
- Slow
- Prohibitive to non-expert routine use

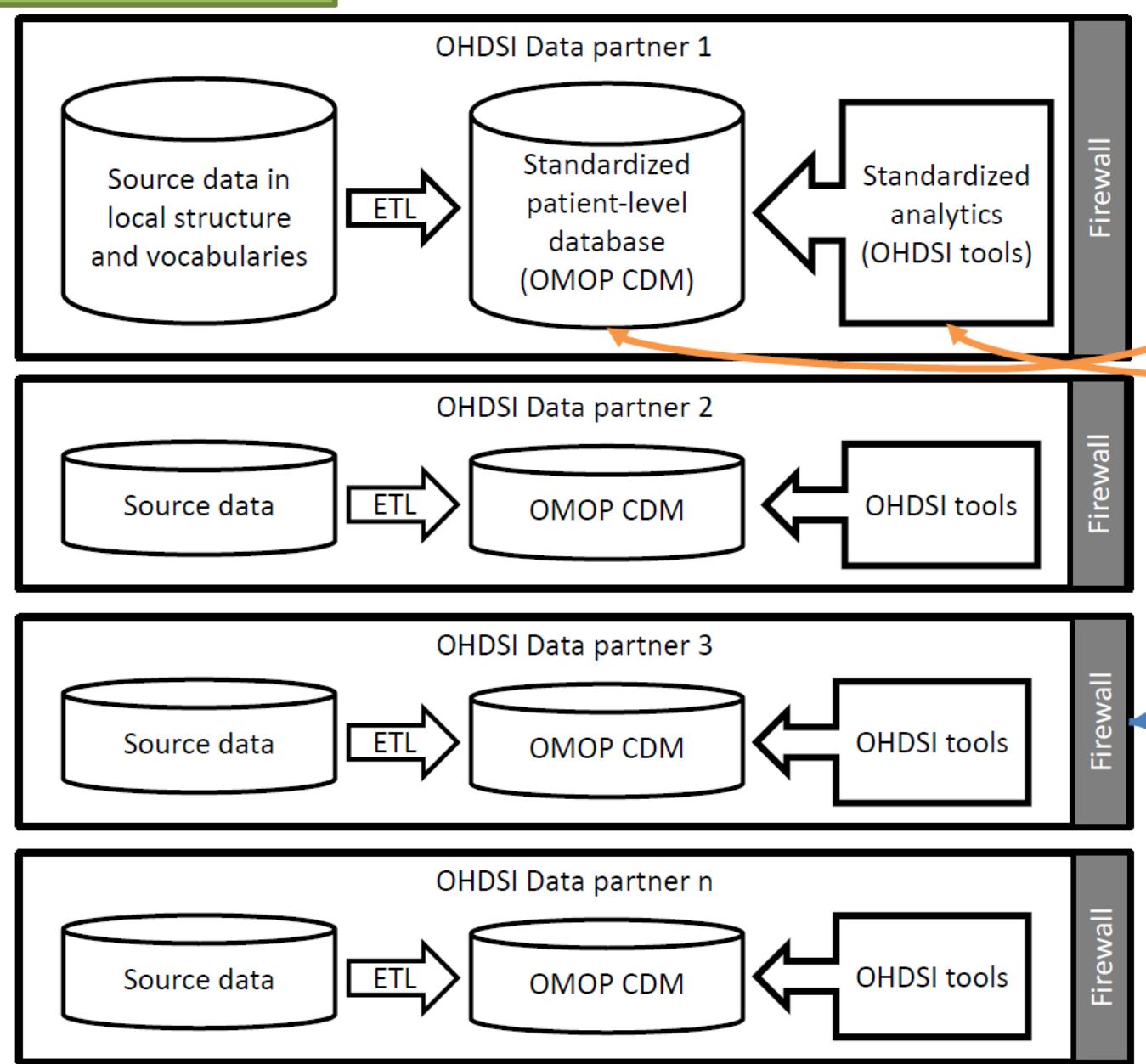
## OMOP CDM



## OHDSI tools



## OHDSI data network



## OHDSI collaborations

Open community data standards (OMOP CDM)

Open source development (OHDSI tools)

Methodological research

Clinical evidence generation

## OHDSI Network studies

Pre-specified protocol with analysis specification

Standardized summary statistics results repository

Collaborative Interpretation

Evidence dissemination



# Health Analytics Data-to-Evidence (HADES)

*Suite of OHDSI R packages for running standardized analytics against OMOP data assets*



## Overview

- R packages that can run against any OMOP database
- Support R packages
  - DatabaseConnector
  - SqlRender
  - ROhdsiWebApi
  - CohortDiagnostics
- Analytical R packages
  - PatientLevelPrediction
  - CohortMethod (comparative effectiveness)
  - FeatureExtraction (characterization)
  - SelfControlledCaseSeries

The screenshot shows the HADES website interface. At the top, there is a logo for HADES (Health Analytics Data-to-Evidence Suite) featuring a stylized orange and yellow square with a white 'X' and a blue arrow pointing right. Below the logo, the word 'HADES' is written in large, bold, black letters, with 'HEALTH ANALYTICS DATA-TO-EVIDENCE SUITE' in smaller text underneath. The main content area is a grid of 16 cards, each representing a different R package. The packages are arranged in four rows of four. Each card includes the package name, a brief description, and a 'Learn more...' link. The packages listed are: CohortMethod, SelfControlledCaseSeries, Cyclops, DatabaseConnector; SelfControlledCohort, EvidenceSynthesis, ParallelLogger, FeatureExtraction; PatientLevelPrediction, EmpiricalCalibration, BigKnn, ROhdsiWebApi; MethodEvaluation, CohortDiagnostics, Hydra, Eunomia; and Andromeda, OhdsiSharing, CirceR.

CohortMethod	SelfControlledCaseSeries	Cyclops	DatabaseConnector
New-user cohort studies using large-scale regression for propensity and outcome models. <a href="#">Learn more...</a>	Self-Controlled Case Series analysis using few or many predictors, includes splines for age and seasonality. <a href="#">Learn more...</a>	Highly efficient implementation of regularized logistic, Poisson and Cox regression. <a href="#">Learn more...</a>	Connect directly to a wide range of database platforms, including SQL Server, Oracle, and PostgreSQL. <a href="#">Learn more...</a>
SelfControlledCohort	EvidenceSynthesis	ParallelLogger	FeatureExtraction
A self-controlled cohort design, where time preceding exposure is used as control. <a href="#">Learn more...</a>	Routines for combining causal effect estimates and study diagnostics across multiple data sites in a distributed study. <a href="#">Learn more...</a>	Support for parallel computation with logging to console, disk, or email. <a href="#">Learn more...</a>	Automatically extract large sets of features for user-specified cohorts using data in the CDM. <a href="#">Learn more...</a>
PatientLevelPrediction	EmpiricalCalibration	BigKnn	ROhdsiWebApi
Build and evaluate predictive models for user-specified outcomes, using a wide array of machine learning algorithms. <a href="#">Learn more...</a>	Use negative control exposure-outcome pairs to profile and calibrate a particular analysis design. <a href="#">Learn more...</a>	A large scale k-nearest neighbor classifier using the Lucene search engine. <a href="#">Learn more...</a>	Interact with OHDSI WebAPI web services. <a href="#">Learn more...</a>
MethodEvaluation	CohortDiagnostics	Hydra	Eunomia
Use real data and established reference sets as well as simulations injected in real data to evaluate the performance of methods. <a href="#">Learn more...</a>	Generate a wide set of diagnostics to evaluate cohort definitions against databases in the CDM. <a href="#">Learn more...</a>	Hydrating package skeletons into executable R study packages based on specifications in JSON format. <a href="#">Learn more...</a>	A standard CDM dataset for testing and demonstration purposes that runs on an embedded SQLite database. <a href="#">Learn more...</a>
Andromeda	OhdsiSharing	CirceR	
Storing very large data objects on a local drive, while still making it possible to manipulate the data in an efficient manner. <a href="#">Learn more...</a>	Securely sharing (large) files between OHDSI collaborators. <a href="#">Learn more...</a>	An R wrapper for Circe, a library for creating cohort definitions, expressing them as JSON, SQL, or Markdown. <a href="#">Learn more...</a>	

<https://ohdsi.github.io/Hades/index.html>



# Data relevance across clinical domains

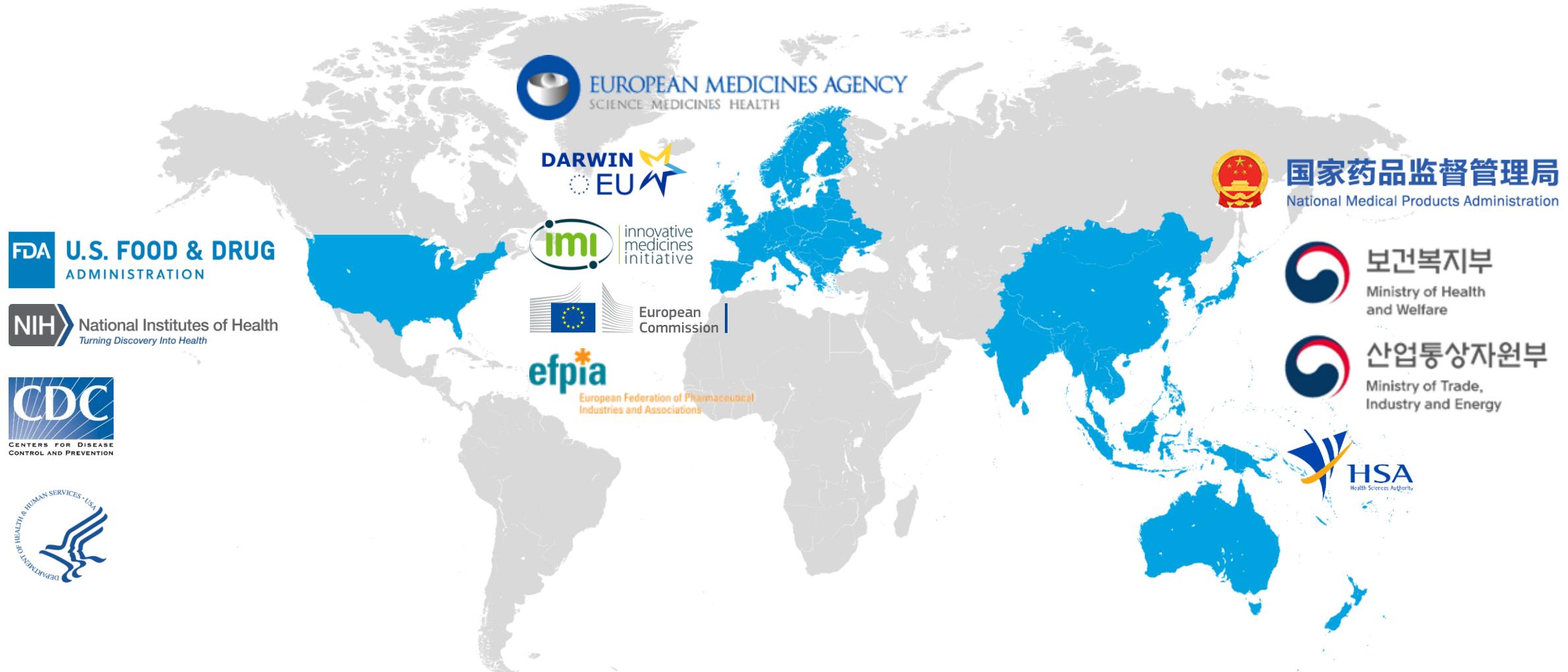
## OMOP Workgroups & OHDSI Phenotype Collaborations

<b>APAC</b> Current Participants: 297 Lead: Mui Van Zandt	<b>ATLAS/WebAPI</b> Current Participants: 253 Lead: Anthony Sena	<b>Clinical Trials</b> Current Participants: 295 Leads: Mike Hamidi, Lin Zhen	<b>CDM</b> Current Participants: 686 Lead: Clair Blacketer	<b>CDM Vocab Subgroup</b> Current Participants: 686 Lead: Michael Kallfelz	<b>Data Network Quality</b> Current Participants: 298 Lead: Clair Blacketer	<b>Dentistry</b> Current Participants: 8 Lead: Robert Koski	<b>Education</b> Current Participants: 136 Lead: Nigel Hughes
<b>HADES</b> Current Participants: 295 Lead: Martijn Schuemie	<b>Health Equity</b> Current Participants: 228 Lead: Jake Gillberg	<b>Latin America</b> Current Participants: 48 Lead: Jose Posada	<b>NLP</b> Current Participants: 444 Lead: Hua Xu	<b>Oncology</b> Current Participants: 328 Lead: Asieh Golozar	<b>Registry</b> Current Participants: 175 Lead: Tina Parciak	<b>Steering Group</b> Current Participants: 82 Lead: Patrick Ryan	<b>Vaccine Vocabulary</b> Current Participants: 79 Lead: Asiyah Lin
<b>Early-Stage Researcher</b> Current Participants: 243 Leads: Faaiyah Arshad, Ross Williams	<b>Eye Care &amp; Vision Research</b> Current Participants: 74 Leads: Sally Baxter, Kerry Goetz	<b>FHIR and OMOP</b> Current Participants: 287 Leads: Jon Duke, Davera Gabriel, Christian Reich	<b>GIS</b> Current Participants: 157 Leads: Robert Miller, Kyle Zollo-Venecek, Andrew Williams	<b>Methods Research</b> Current Participants: 379 Leads: Martijn Schuemie, Marc Suchard	<b>Perinatal &amp; Reproductive Health Group</b> Current Participants: 30 Leads: Alison Callahan et al.	<b>Psychiatry</b> Current Participants: 132 Leads: Dmitry Dymshyts, Andrew Williams	<b>Surgery &amp; Perioperative Medicine</b> Current Participants: 42 Leads: Jenny Lane, Evan Minty
	<b>Medical Imaging</b> Current Participants: 155 Leads: Paul Nagy, Seng Chan You	<b>Medical Devices</b> Current Participants: 141 Leads: Vojtech Huser, Asiyah Lin	<b>Open-Source Community</b> Current Participants: 145 Leads: Adam Black, Paul Nagy	<b>Patient-Level Prediction</b> Current Participants: 89 Leads: Jenna Reps, Ross Williams	<b>Healthcare Systems</b> Current Participants: 471 Lead: Melanie Philofsky	<b>Phenotype</b> Current Participants: 310 Lead: Gowtham Rao	

- [Type 2 Diabetes Mellitus](#)
- [Type 1 Diabetes Mellitus](#)
- [Atrial Fibrillation](#)
- [Multiple Myeloma](#)
- [Alzheimer's Disease](#)
- [Hemorrhagic Events](#)
- [Neutropenia](#)
- [Parkinson's Disease and Parkinsonism](#)
- [Attention Deficit Hyperactivity Disorder](#)
- [Hypertension](#)
- [Acute Myocardial Infarction](#)
- [Heart Failure](#)
- [Cardiomyopathy](#)
- [Multiple Sclerosis](#)
- [Hidradenitis Suppurativa](#)
- [Anaphylaxis](#)
- [Depression](#)
- [Non-Small-Cell Lung Cancer](#)
- [Drug-Induced Liver Injury](#)
- [Severe Visual Impairment And Blindness](#)
- [Suicide Attempts](#)
- [Kidney Stones](#)
- [Delirium](#)
- [Systemic Lupus Erythematosus](#)
- [Triple Negative Breast Cancer](#)
- [Pulmonary Hypertension](#)
- [Prostate Cancer](#)
- [HIV](#)



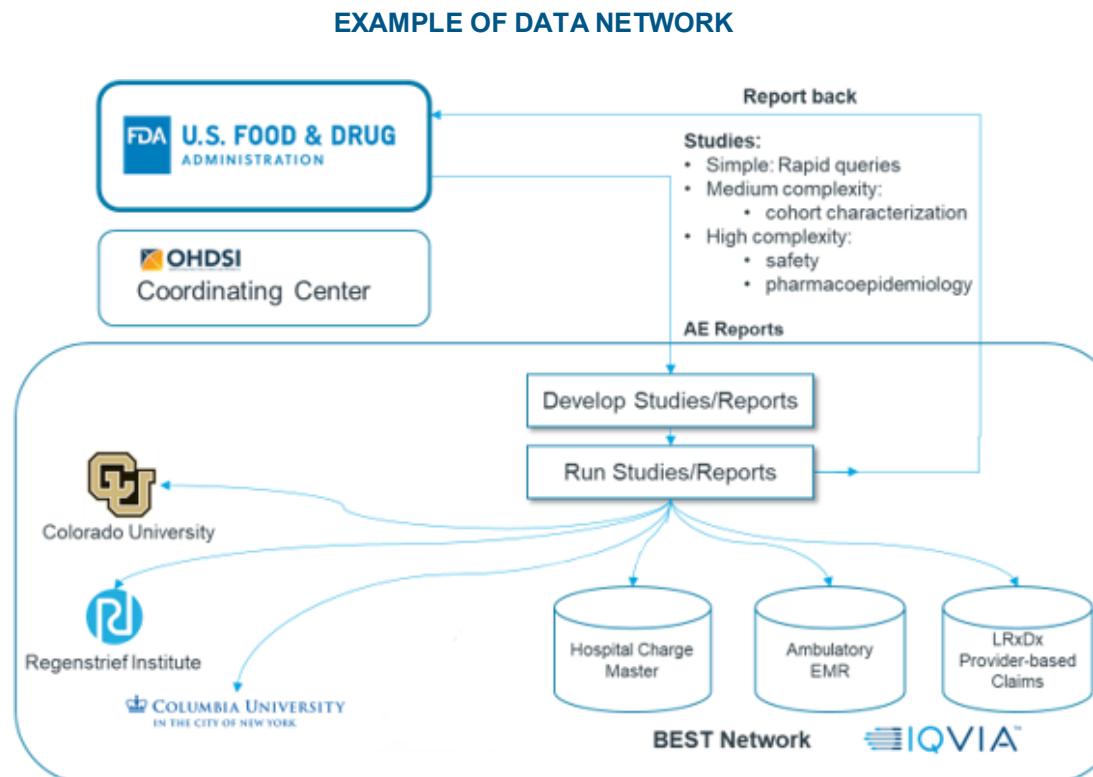
# Global OHDSI Adoptions





# FDA BEST – Overview

Premier, multi-center research collaborative driving large scale health analytics research



## Network Overview

- Started in September 2017
- Today's largest distributed network of clinical data
- Collaborative research model, guided by efforts across the OHDSI community and US FDA
- Iterative sponsored studies facilitated by IQVIA and the global network of data partners



## Benefits to Participating Sites

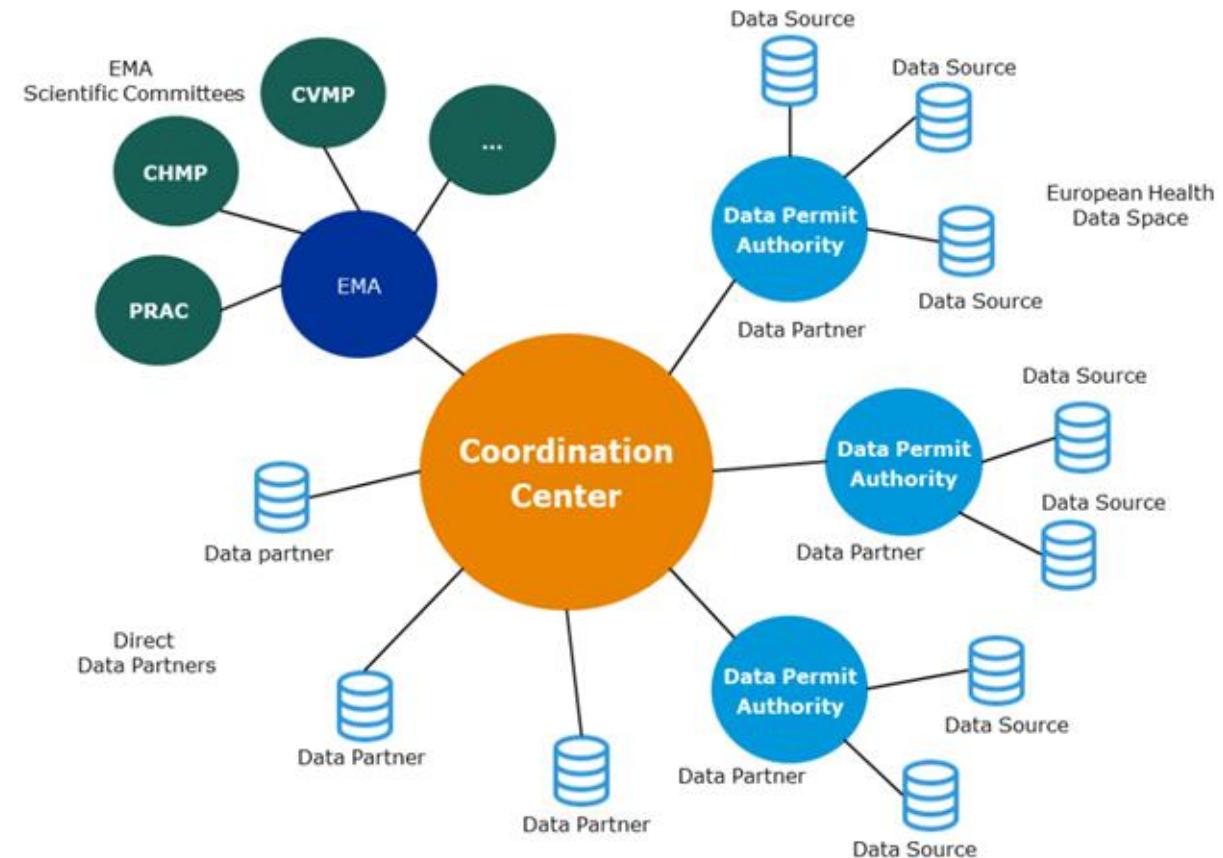
- Access to large, diverse patient populations
- Maintain direct control of your site's clinical data, share only aggregate data
- Access to IQVIA data enrichment programs to enhance site data (e.g. NLP tools, linkage services)
- Ability for researchers to externally validate single-center findings



# Data Analysis and Real-World Interrogation Network (DARWIN EU®)

A paradigm shift for the use of real-world health data for regulatory purpose in the EU

DARWIN EU® is a federated network of data, expertise, and services that supports better decision-making throughout the product lifecycle by generating reliable evidence from real world healthcare data





# China Government's Guides on RWE & RWD

*From Center for Drug Evaluation (CDE), National Medical Products Administration (NMPA)*

- **1<sup>st</sup> guide** was released in Jan 2020, introducing the definition, data source requirement, design, and evaluation of using RWE for drug effectiveness study and safety monitoring.
- **2<sup>nd</sup> guide** was released in Aug 2020, focusing on the details and importance of the source, safety, curation, quality assurance and maintenance of RWD, so that reliable RWE could be produced

国家药品监督管理局药品审评中心  
NMPA CDE CENTER FOR DRUG EVALUATION, NMPA  
此页面上的内容需要较新版本的 Adobe Flash Player.  
Get ADOBE® FLASH® PLAYER

当前位置：新闻中心>>工作动态>>通知公告>>新闻正文

关于公开征求《用于产生真实世界证据的真实世界数据指导原则（征求意见稿）》意见的通知  
发布日期：20200803

为进一步指导和规范申办者利用真实世界数据生成真实世界证据支持药物研发，我中心组织起草了《用于产生真实世界证据的真实世界数据指导原则（征求意见稿）》，现在中心网站予以公示，以广泛听取各界意见和建议，欢迎各界提出宝贵意见和建议，并请及时反馈给我们。

征求意见时限为自发布之日起2个月。

您的反馈意见请发到以下联系人的邮箱：

联系人：高丽丽、赵骏  
联系方式：gaoll@cde.org.cn, zhaojun@cde.org.cn

感谢您的参与和大力支持。

国家药品监督管理局药品审评中心  
2020年8月3日

附件 1：	《用于产生真实世界证据的真实世界数据指导原则（征求意见稿）》.docx
附件 2：	《用于产生真实世界证据的真实世界数据指导原则（征求意见稿）》起草说明.doc



# China Government's Guides on RWE & RWD

*CDM & OHDSI Citations in the 2<sup>nd</sup> Guide, Section 4 – Real World Data Curation*

## CDM Introduction in Guide:

- Under multidisciplinary collaboration, CDM was created with standardized structure, format and vocabulary, to achieve multi-center data integration and collaboration.

## References in Guide:

- EMA. A Common Data Model for Europe – Why? Which? How?  
<https://www.ema.europa.eu/en/events/common-data-model-europe-why-which-how>
- OHDSI – Observational Health Data Sciences and Informatics, <https://www.ohdsi.org>

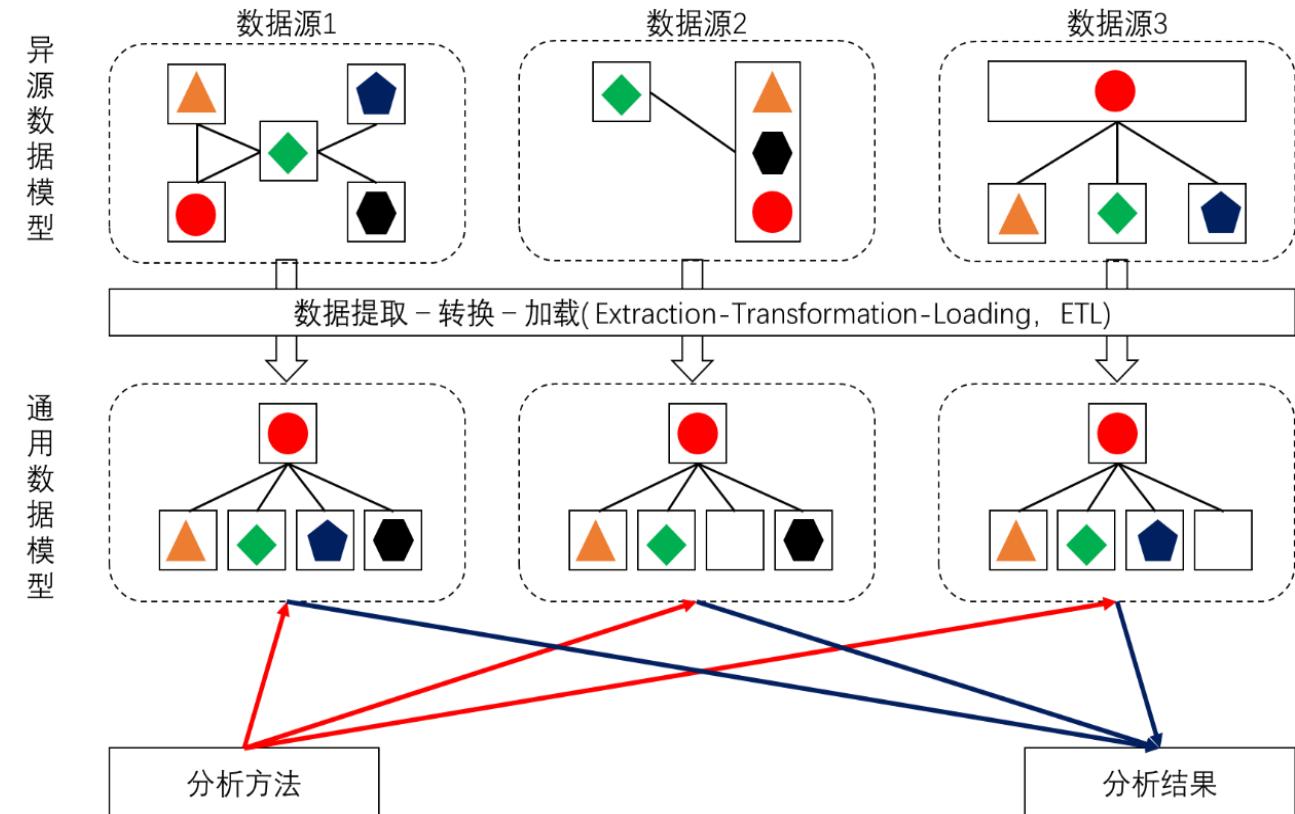
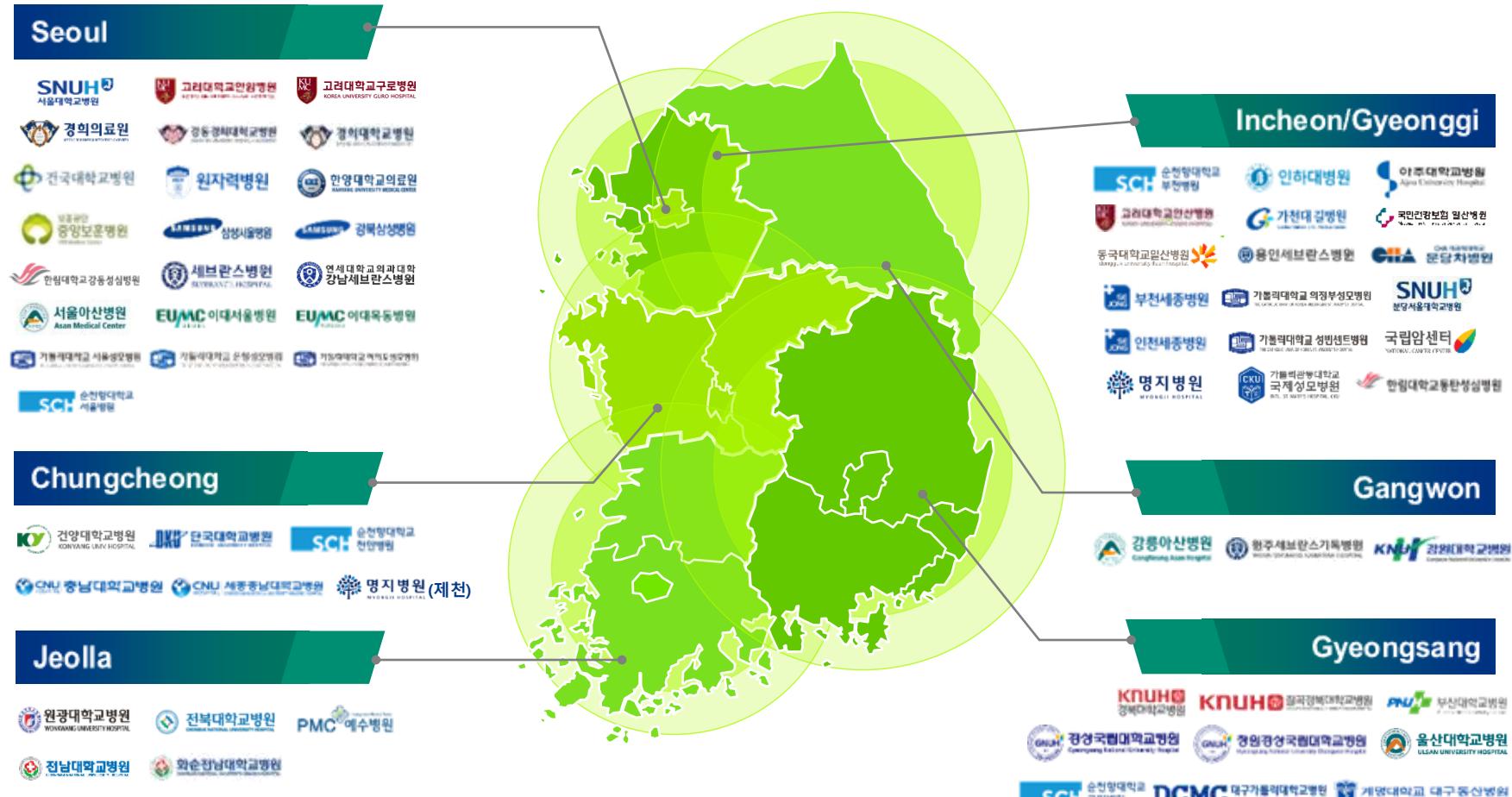


Fig. 2 in Guide – Diagram on Converting Source Data to CDM



# Korean Government Initiatives



Federated OMOP network of 62 hospitals and >76M patients



## Participation from secondary and tertiary hospitals nationwide



Good representation of each of the provinces of South Korea



Funded by series of grants from  
the Korean government

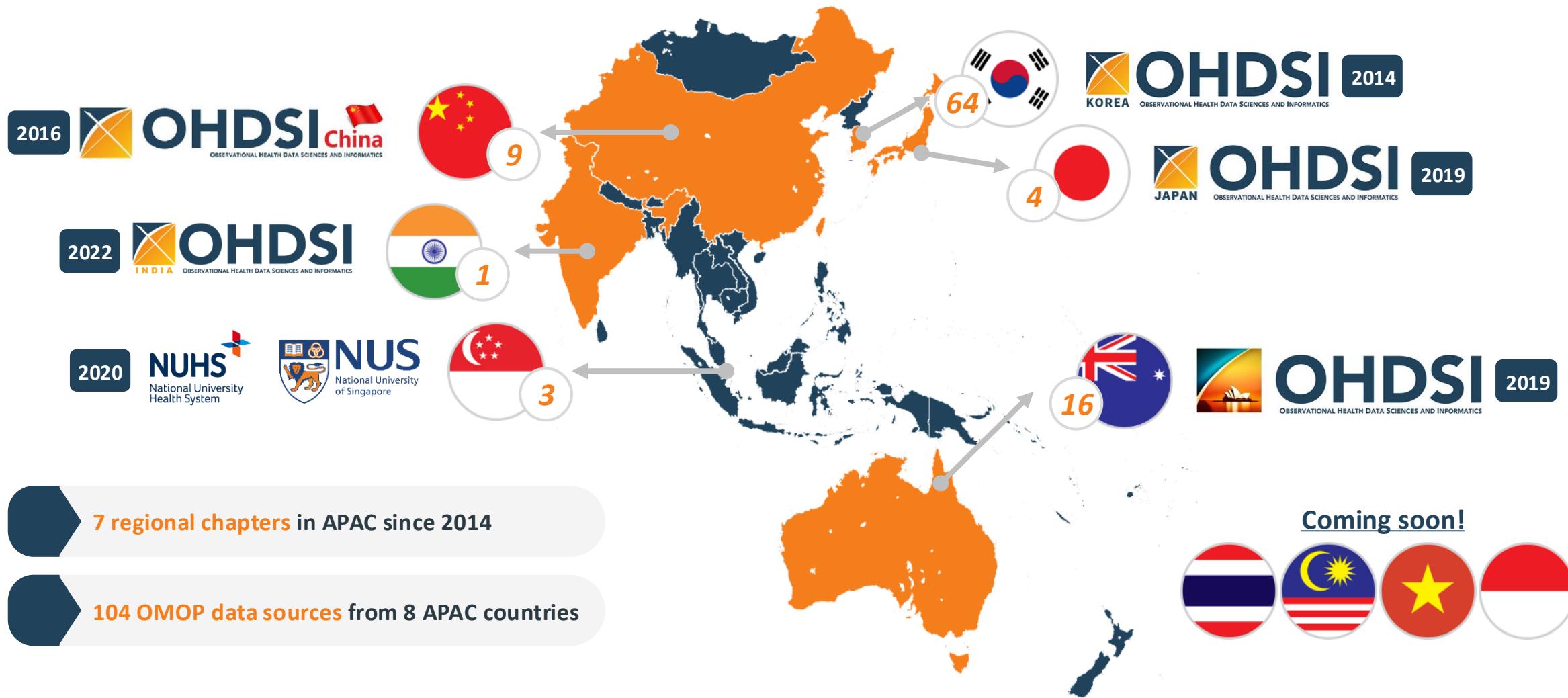


## Foster collaborative research and evidence-sharing ecosystem





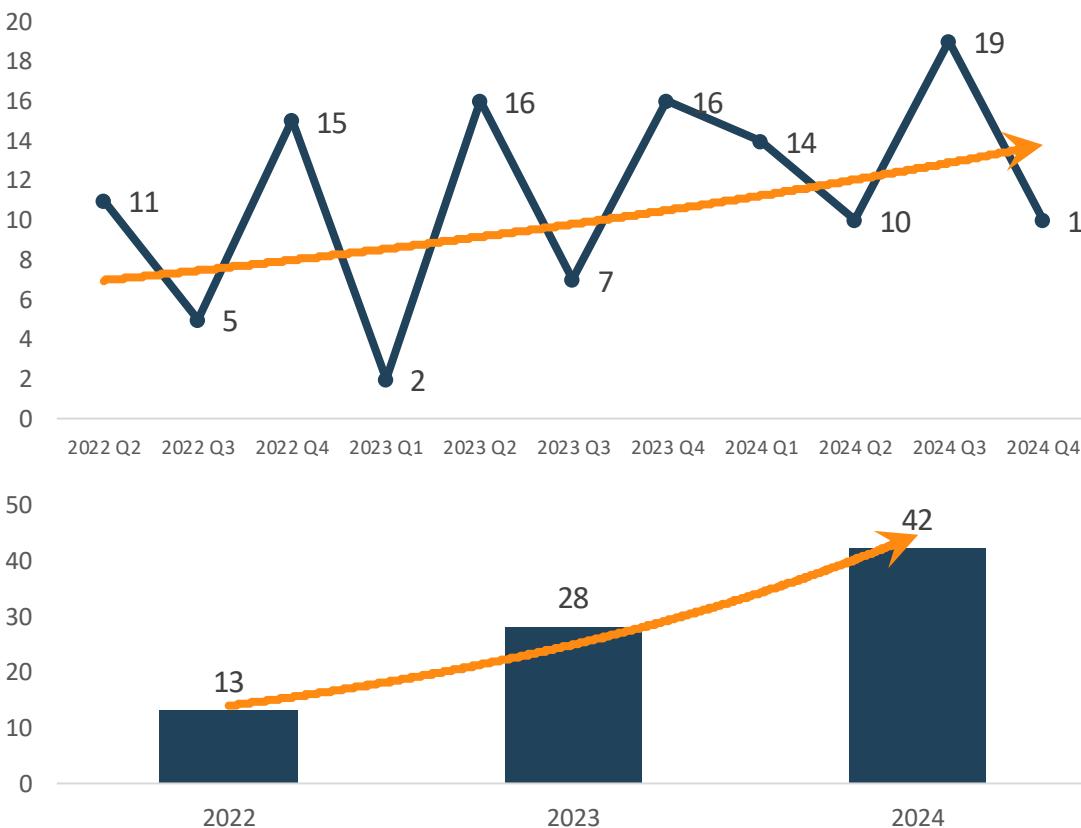
# OHDSI APAC





# APAC Studies

**Increasing trend of publications and abstract submissions from APAC**



JAMA Network **Open**



Original Investigation | Pharmacy and Clinical Pharmacology

## Ranitidine Use and Incident Cancer in a Multinational Cohort

Seng Chan You, MD; Seung In Seo, MD; Thomas Falconer, MSc; Chen Yanover, PhD; Talita Duarte-Salles, PhD; Sarah Seager, BA; Jose D. Posada, PhD; Nigam H. Shah, PhD; Phung-Anh Nguyen, PhD; Yeesuk Kim, MD; Jason C. Hsu, PhD; Mui Van Zandt, BS; Min-Huel Hsu, MD; Hang Lak Lee, MD; Heejoo Ko, MD; Woon Geon Shin, MD;

Research

JAMA Psychiatry | Original Investigation

## Rates of Antipsychotic Drug Prescribing Among People Living With Dementia During the COVID-19 Pandemic

Hao Luo, PhD; Wallis C. Y. Lau, PhD; Yi Chai, PhD; Carmen Olga Torre, MSc; Robert Howard, MD; Kathy Y. Liu, PhD; Xiaoyu Lin, MSc; Can Yin, MSc; Stephen Fortin, PharmD; David M. Kern, PhD; Dong Yun Lee, MD;



Psychotropic drug prescribing before and during the COVID-19 pandemic among people with depressive and anxiety disorders: a multinational network study

Hao Luo\*, Yi Chai\*, Sijia Li, Wallis CY Lau, Carmen Olga Torre, Joseph Hayes, Ivan CH Lam, Xiaoyu Lin, Can Yin, Stephen Fortin, Dave M Kern, Dong Yun Lee, Rae Woong Park, Jae-Won Jang, Celine SL Chui, Jing Li, Sarah Seager, Kenneth K CM Man, Ian CK Wong

**100+ publications authored by APAC collaborators including numerous papers in high-impact journals**



# 2025 APAC Study

***Association Between Fasting Plasma Glucose Levels and Annual Hospitalization Days*** (Fudan University)

***Gastrointestinal Risk of GLP-1 Receptor Agonists versus SGLT-2 and DPP-4 Inhibitors in Type 2 Diabetes*** (Peking University)

***Studying the Disease Trajectory of Type 2 Diabetes with Transformer-based Model*** (University of Science and Technology of China)

- Studies initiated and led by China
- Ongoing work to align research questions to OHDSI's standardized methods and analytical frameworks



# Titans in APAC

## 2023 Titan Awards 2024 Titan Awards



Nicole Pratt



[www.ohdsi.org](http://www.ohdsi.org) Community Leadership #JoinTheJourney



Cindy Cai



Jung Ho Kim



Jack Janetzki

## 2023 Titan Awards



Gyeol Song



[www.ohdsi.org](http://www.ohdsi.org) Community Support #JoinTheJourney



Natthawut 'Max' Adulyanukosol



[www.ohdsi.org](http://www.ohdsi.org) Community Collaboration #JoinTheJourney



# Summary

1

## Open Source

CDM, tools, methods, and documentation all publicly available

2

## Standardization

Standard CDM, vocabulary/ontology, tools, methods, data quality, and documentation

3

## Research Community

Large research community with multiple stakeholders and disciplines

4

## Multi-country/multi-center research

Large scale research using standardized tools and methods



**OHDSI**

OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

Thank you!