2024 OHDSI APAC Lightning Talk

Feasibility of Integrating DICOM Headers into the OMOP Medical Imaging CDM
: A Pilot Study Using Chest CT Data

Kyulee Jeon, B.S.
Department of Biomedical Systems Informatics,
Yonsei University College of Medicine, Seoul, Republic of Korea

Medical Imaging CDM (MI-CDM)

- **The OHDSI Medical Image Workgroup** proposed two new tables to the OMOP CDM, the medical imaging extension model

Journal of Imaging Informatics in Medicine (2024) 37:899–908
https://doi.org/10.1007/s10278-024-00982-6

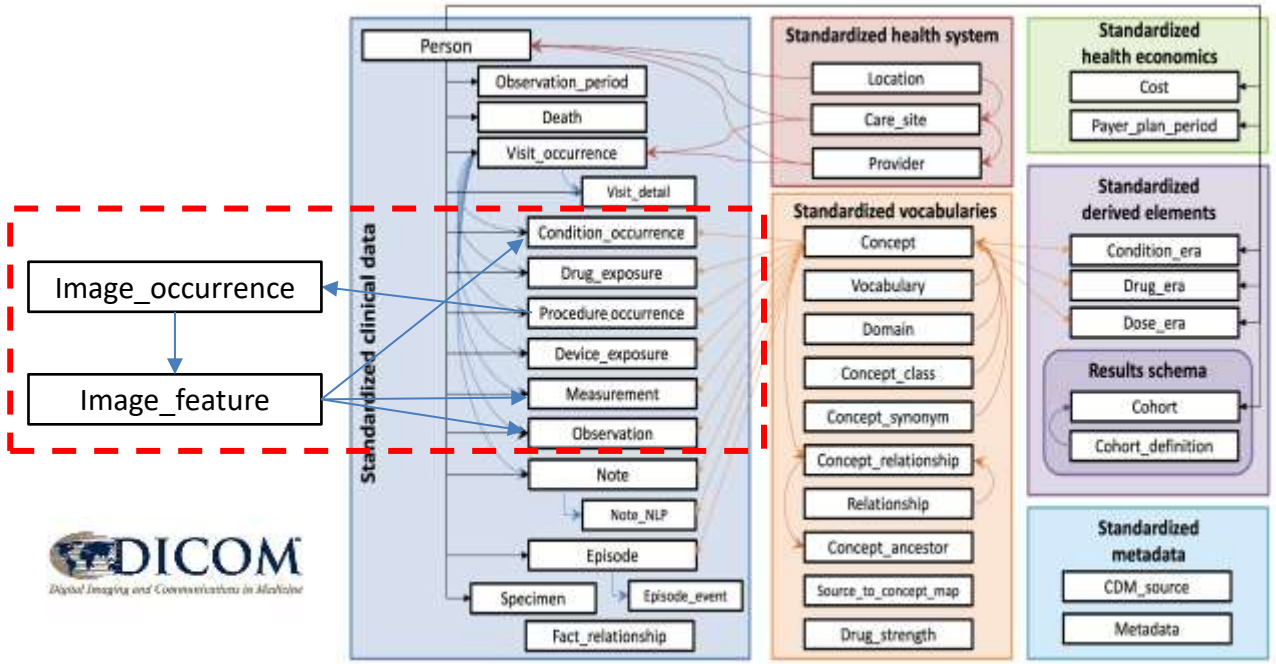
Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension

Woo Yeon Park[1] · Kyulee Jeon[2,3] · Teri Sippel Schmidt[1] · Haridimos Kondylakis[4] · Tarik Alkasab[5] · Blake E. Dewey[6] · Seng Chan You[2,3] · Paul Nagy[1]

Received: 4 September 2023 / Revised: 10 November 2023 / Accepted: 14 November 2023 / Published online: 5 February 2024

Abstract
The rapid growth of artificial intelligence (AI) and deep learning techniques require access to large inter-institutional cohorts of data to enable the development of robust models, e.g., targeting the identification of disease biomarkers and quantifying disease progression and treatment efficacy. The Observational Medical Outcomes Partnership Common Data Model (OMOP CDM) has been designed to accommodate a harmonized representation of observational healthcare data. This study proposes the Medical Imaging CDM (MI-CDM) extension, adding two new tables and two vocabularies to the OMOP CDM to address the structural and semantic requirements to support imaging research. The tables provide the capabilities of linking DICOM data sources as well as tracking the provenance of imaging features derived from those images. The implementation of the extension enables phenotype definitions using imaging features and expanding standardized computable imaging biomarkers. This proposal offers a comprehensive and unified approach for conducting imaging research and outcome studies utilizing imaging features.

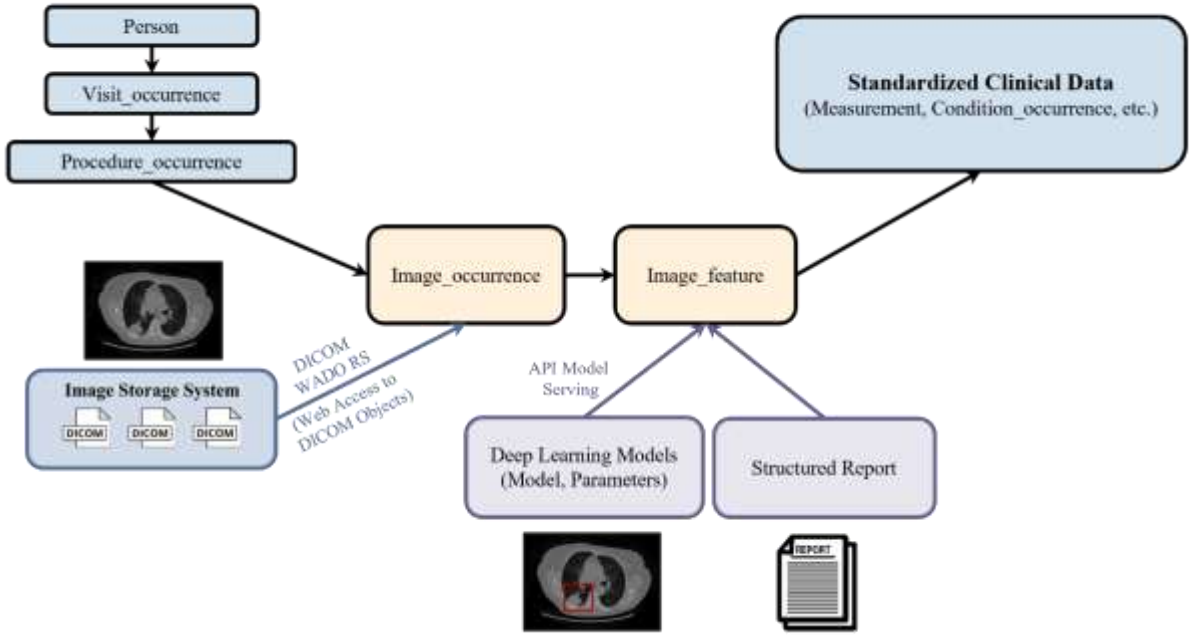
Keywords Data collection [MeSH] · Data standardization · Observational research · Data integration · Multimodal data analysis



Park, et al. “Development of Medical Imaging Data Standardization for Imaging-Based Observational Research: OMOP Common Data Model Extension.” Journal of Imaging Informatics in Medicine, February 5, 2024

Imaging Goals with OHDSI

- Bring **features** derived from medical images into data model while **maintaining provenance**
  - E.g. Lung nodule's size detected by AI
  - E.g. CT slice thickness recorded in DICOM
- Perform **cohort definitions** in OHDSI for **medical imaging studies**
  - E.g. Find the Chest CT Scans with a slice thickness of <2.5 mm for patients ultimately diagnosed with lung cancer



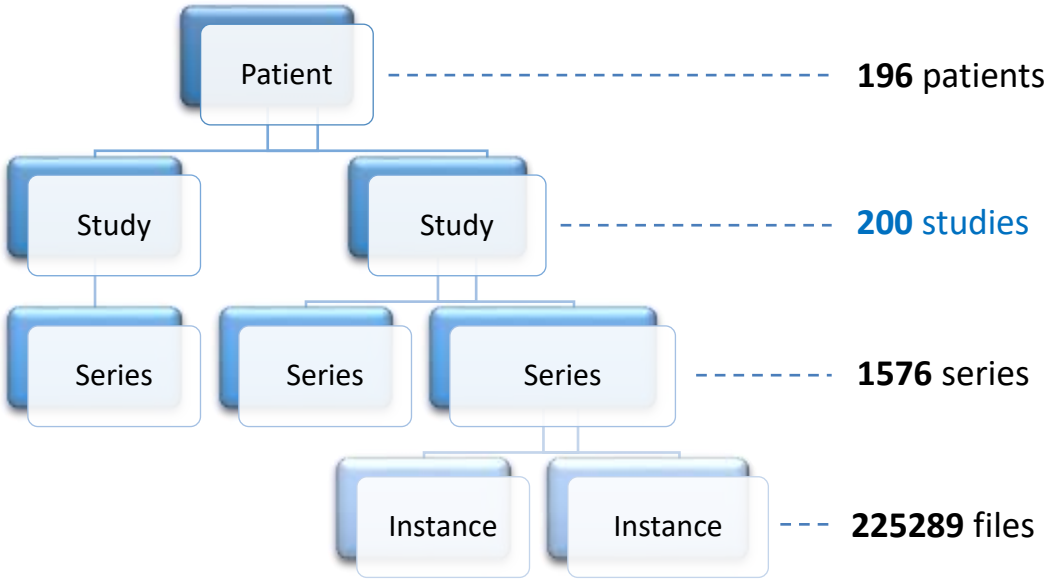
[Fig 1. Framework of Medical Imaging CDM]

Pilot Study: Objectives

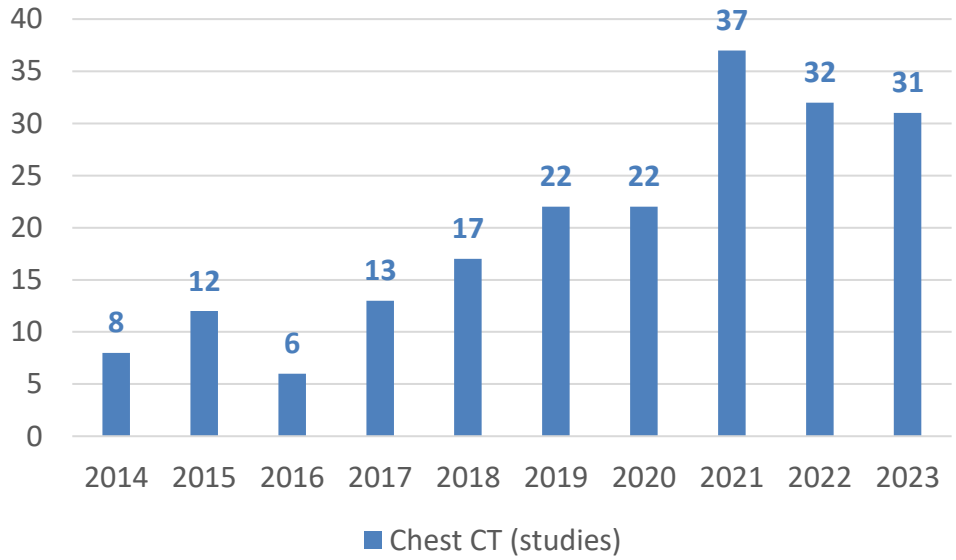
- Implement the MI-CDM **using real-world imaging data** and evaluate the feasibility of **integrating DICOM headers** into the MI-CDM framework

Data Source

- Severance Hospital, Seoul, South Korea
  - **200 chest CT scans** from 196 lung cancer patients (diagnosed 2014-2023)



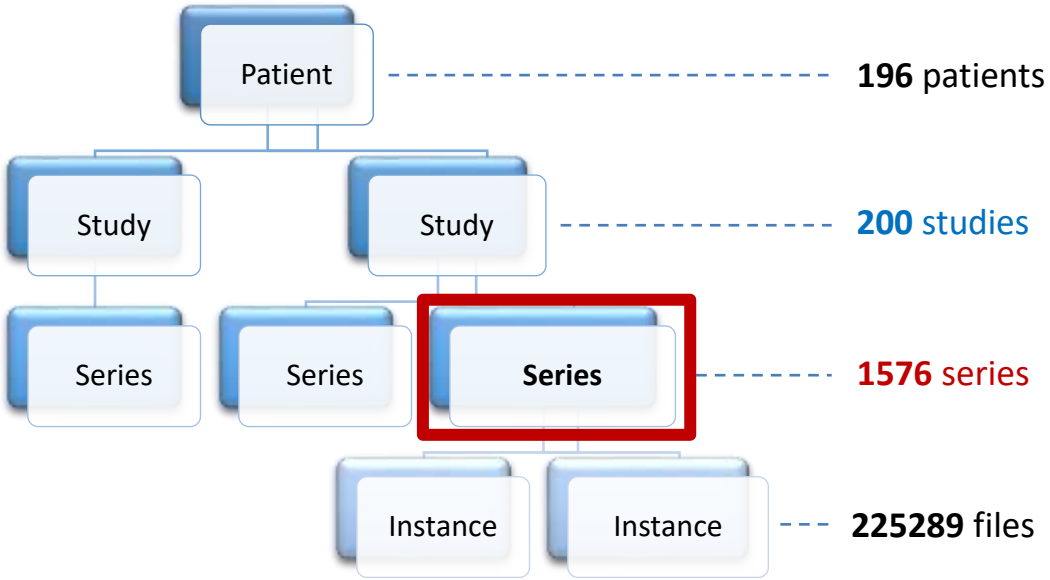
[Fig 2. Hierarchical Structure of the DICOM Dataset]



[Fig 3. Yearly Distribution of Chest CT Dataset (2014-2023)]

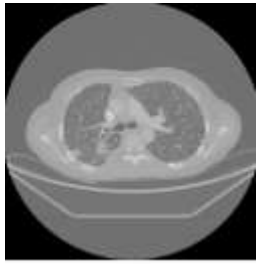
Data Source

- Severance Hospital, Seoul, South Korea
  - 200 chest CT scans from 196 lung cancer patients (diagnosed 2014-2023)



[Fig 2. Hierarchical Structure of the DICOM Dataset]

(Example: Different series within same study for single patient at same time)

(PAT123, 2020.10.23)	Series #1	Series #2
SliceThickness	800	2.5
KVP	120	100
ScanOptions	SCOUT MODE	HELICAL MODE
ContrastBolusVolume	100	94
Pixel Image (Example)		

DICOM Dataset

- DICOM Metadata
  - **Descriptive information** about the patient, study, and imaging parameters.
  - Pixel values depend on metadata such as **acquisition parameters**.
- Link DICOM metadata to OMOP for imaging data discovery, retrieval, and analysis.

DICOM®
Digital Imaging and Communications in Medicine

Tag	Attribute Name	Value (example)
(0010,0040)	Patient's Sex	M
(0010,1010)	Patient's Age	067Y
(0008,0020)	Study Date	20190101
(0008,0060)	Modality	CT
(0018,0015)	Body Part Examined	CHEST
(0008,0070)	Manufacturer	Siemens Healthineers
(0008,1090)	Manufacturer's Model Name	SOMATOM Force
(0018,0050)	Slice Thickness	1
(0018,0060)	KVP	100
(0018,1041)	Contrast/Bolus Volume	0
(0018,9345)	CTDIvol	4.13998747826087 (mGy)

[Table 1. DICOM Metadata Example for a Chest CT Image]

DICOM to OMOP ETL Process with an Example

DICOM Header Entry

- Patient ID: PAT123
- Study Date: 20230615
- Modality: CT
- Body Part: CHEST
- KVp: 120
- Slice Thickness: 3mm

Processing Steps:

1. Structural Processing

Series-level Data
- Modality: CT
- Body Part: CHEST

Instance-level Data
- KVp: 120
- Slice Thickness: 3mm

2. Semantic Mapping

Tags to Customed Concept IDs
- "KVp" → 2128000358 (DICOM Tag)
- "Slice Thickness" → 2128000357 (DICOM Tag)

Values to Standard Concept IDs
- CHEST → 37303885 (SNOMED)

Values to Customed Concept IDs
- CT → 2128008068 (DICOM Value Set)

3. Value Processing

Tag Mapping Results
"KVp", "Slice Thickness" → measurement_concept_id

Numeric Values
120, 3 → value_as_number

Categorical Values
CT, CHEST → value_as_concept_id

OMOP Tables:

Image_occurrence Table
- image_occurrence_id: 1
- person_id: 12345
- procedure_occurrence_id: ###
- image_occurrence_date: 2023-06-15
- modality_concept_id: 4107854
- anatomic_site_concept_id: 3703885
- local_path: /yuhs/study1/series1/

Measurement Table
- person_id: 12345
- measurement_id: 11
- measurement_date: 2023-06-15
- measurement_concept_id: 2128000358
- value_as_number: 120

- person_id: 12345
- measurement_id: 12
- measurement_date: 2023-06-15
- measurement_concept_id: 2128000357
- value_as_number: 3

Image_feature Table
- person_id: 12345
- image_occurrence_id: 1
- image_feature_event_field_concept_id: 1147330
- image_feature_event_id: 11
- image_feature_concept_id: 2128000358

- person_id: 12345
- image_occurrence_id: 1
- image_feature_event_field_concept_id: 1147330
- image_feature_event_id: 12
- image_feature_concept_id: 2128000357

Example: A DICOM Parameter (KVp) into OMOP

DICOM Header Entry

- Patient ID: PAT123
- Study Date: 20230615
- Modality: CT
- Body Part: CHEST
- KVp: 120
- Slice Thickness: 3mm

Processing Steps:

1. Structural Processing

Series-level Data
- Modality: CT
- Body Part: CHEST

Instance-level Data
- KVp: 120
- Slice Thickness: 3mm

2. Semantic Mapping

Tags to Customed Concept IDs
- "KVp" → 2128000358 (DICOM Tag)
- "Slice Thickness" → 2128000357 (DICOM Tag)

Values to Standard Concept IDs
- CHEST → 37303885 (SNOMED)

Values to Customed Concept IDs
- CT → 2128008068 (DICOM Value Set)

3. Value Processing

Tag Mapping Results
"KVp", "Slice Thickness" → measurement_concept_id

Numeric Values
120, 3 → value_as_number

Categorical Values
CT, CHEST → value_as_concept_id

OMOP Tables:

Image_occurrence Table
- image_occurrence_id: 1
- person_id: 12345
- procedure_occurrence_id: ###
- image_occurrence_date: 2023-06-15
- modality_concept_id: 4107854
- anatomic_site_concept_id: 3703885
- local_path: /yuhs/study1/series1/

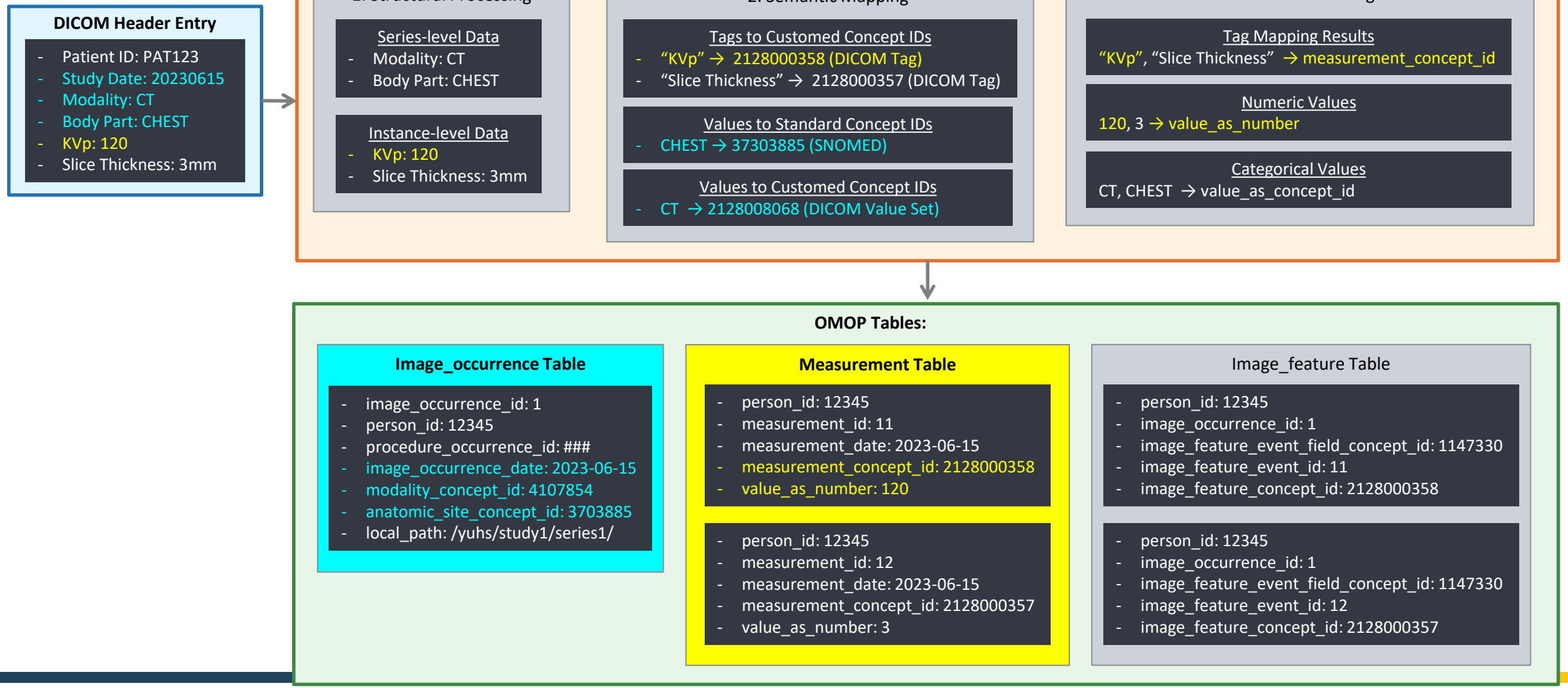
Measurement Table
- person_id: 12345
- measurement_id: 11
- measurement_date: 2023-06-15
- measurement_concept_id: 2128000358
- value_as_number: 120

- person_id: 12345
- measurement_id: 12
- measurement_date: 2023-06-15
- measurement_concept_id: 2128000357
- value_as_number: 3

Image_feature Table
- person_id: 12345
- image_occurrence_id: 1
- image_feature_event_field_concept_id: 1147330
- image_feature_event_id: 11
- image_feature_concept_id: 2128000358

- person_id: 12345
- image_occurrence_id: 1
- image_feature_event_field_concept_id: 1147330
- image_feature_event_id: 12
- image_feature_concept_id: 2128000357

Example: Series-level Imaging Provenance



DICOM to OMOP ETL Process with an Example

DICOM Header Entry

- Patient ID: PAT123
- Study Date: 20230615
- Modality: CT
- Body Part: CHEST
- KVp: 120
- Slice Thickness: 3mm

Processing Steps:

1. Structural Processing

Series-level Data
- Modality: CT
- Body Part: CHEST

Instance-level Data
- KVp: 120
- Slice Thickness: 3mm

2. Semantic Mapping

Tags to Customed Concept IDs
- "KVp" → 2128000358 (DICOM Tag)
- "Slice Thickness" → 2128000357 (DICOM Tag)

Values to Standard Concept IDs
- CHEST → 37303885 (SNOMED)

Values to Customed Concept IDs
- CT → 2128008068 (DICOM Value Set)

3. Value Processing

Tag Mapping Results
"KVp", "Slice Thickness" → measurement_concept_id

Numeric Values
120, 3 → value_as_number

Categorical Values
CT, CHEST → value_as_concept_id

OMOP Tables:

Image_occurrence Table

- image_occurrence_id: 1
- person_id: 12345
- procedure_occurrence_id: ###
- image_occurrence_date: 2023-06-15
- modality_concept_id: 4107854
- anatomic_site_concept_id: 3703885
- local_path: /yuhs/study1/series1/

Measurement Table

- person_id: 12345
- measurement_id: 11
- measurement_date: 2023-06-15
- measurement_concept_id: 2128000358
- value_as_number: 120

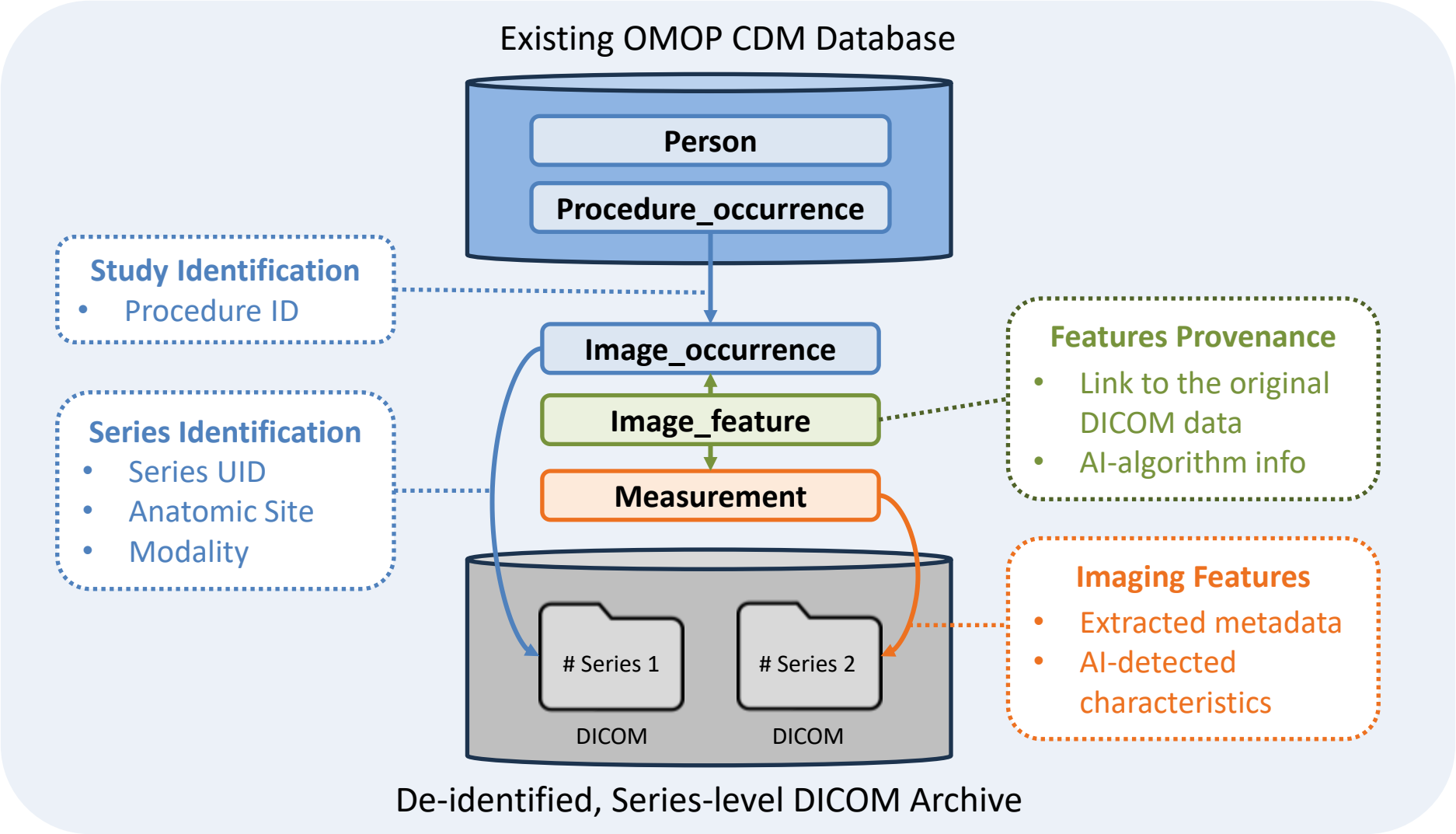
- person_id: 12345
- measurement_id: 12
- measurement_date: 2023-06-15
- measurement_concept_id: 2128000357
- value_as_number: 3

Image_feature Table

- person_id: 12345
- image_occurrence_id: 1
- image_feature_event_field_concept_id: 1147330
- image_feature_event_id: 11
- image_feature_concept_id: 2128000358

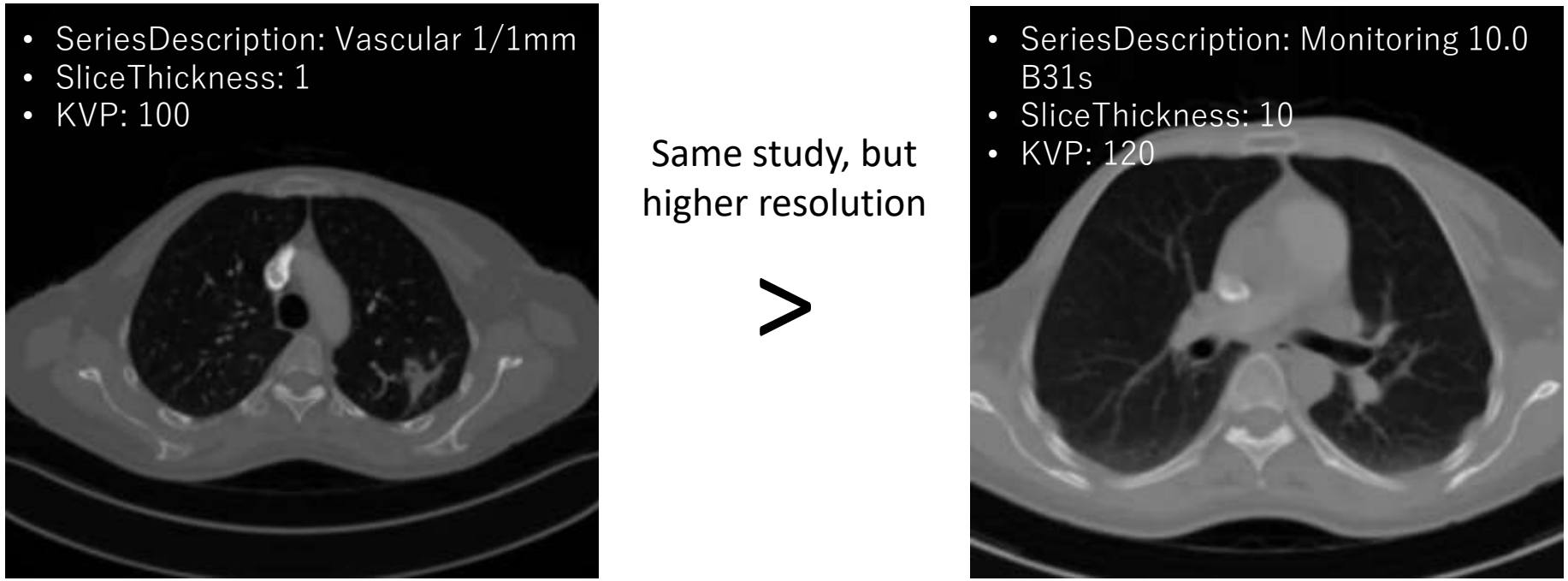
- person_id: 12345
- image_occurrence_id: 1
- image_feature_event_field_concept_id: 1147330
- image_feature_event_id: 12
- image_feature_concept_id: 2128000357

MI-CDM Workflow



Use Case Example

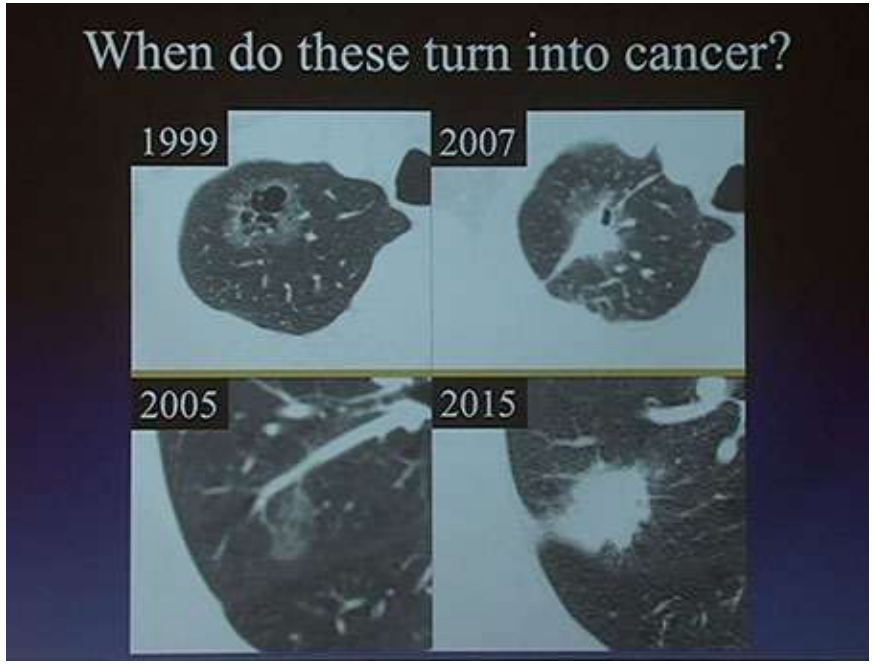
- Define a cohort of lung cancer patients who underwent chest CT scans within 7 days of surgery.
- Retrieve imaging series with a slice thickness ≤ 1 mm.



[Fig 4. Different series within the same chest CT study (same patient, same day)

Conclusion

- Enhanced **imaging data retrieval** based on patient and imaging parameters.
- Incorporating **pixel phenotypes** into the OMOP CDM broadens the scope of capturing the patient journey.



Acknowledgement

Special thanks to the OHDSI Medical Image Workgroup:

Jen Woo Yeon Park, M.S.,

Teri Sippel Schmidt, M.S.,

Blake Dewey, Ph.D.,

Paul Nagy, Ph.D.,

Seng Chan You, M.D., Ph.D.

Thank you!

Join the OHDSI community and Medical Imaging Working Group

https://www.ohdsi.org/join-the-journey/

Additional information on OMOP CDM Medical Imaging extension can be found here.

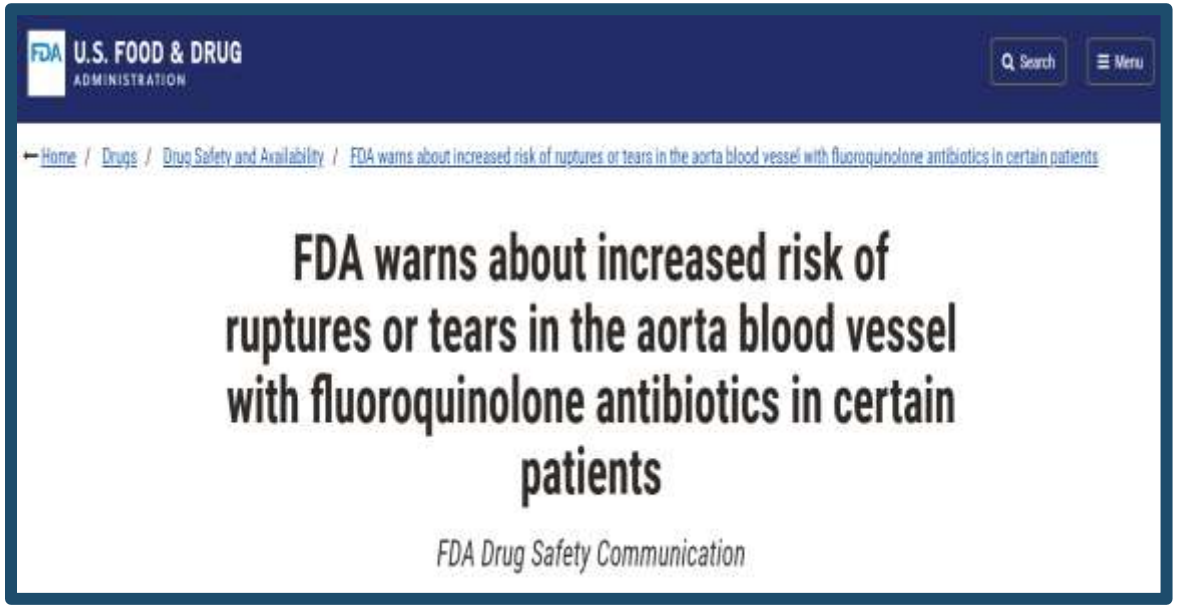
Contact: jklee320@yuhs.ac, chandryou@yuhs.ac

Risk of aortic aneurysm or dissection following use of fluoroquinolones: a multinational network cohort study

Jack L Janetzki[1,*], Jung Ho Kim,[2,*] Nicole Pratt,[1] and Seng Chan You, [3,4] on behalf to the 2023 OHDSI SOS Challenge Fluoroquinolone team

Background

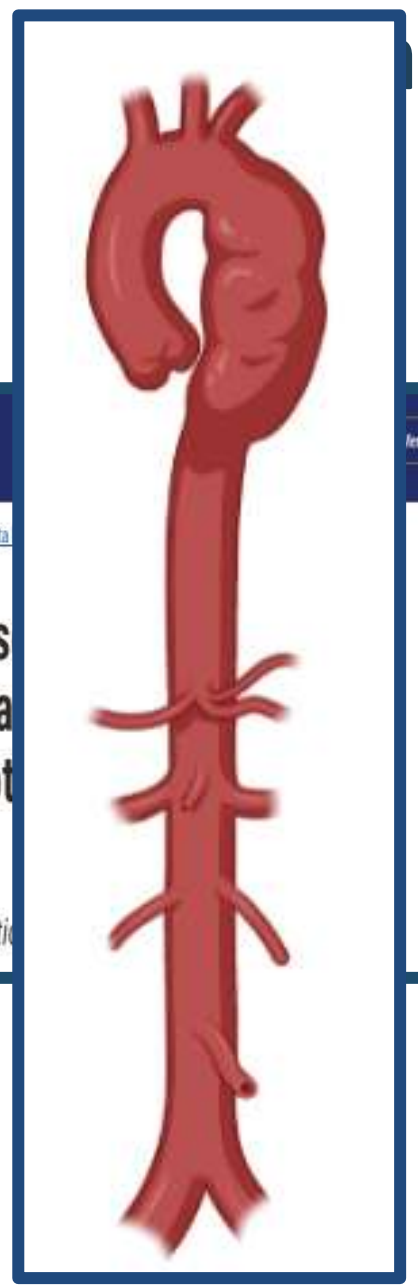
December 2018:

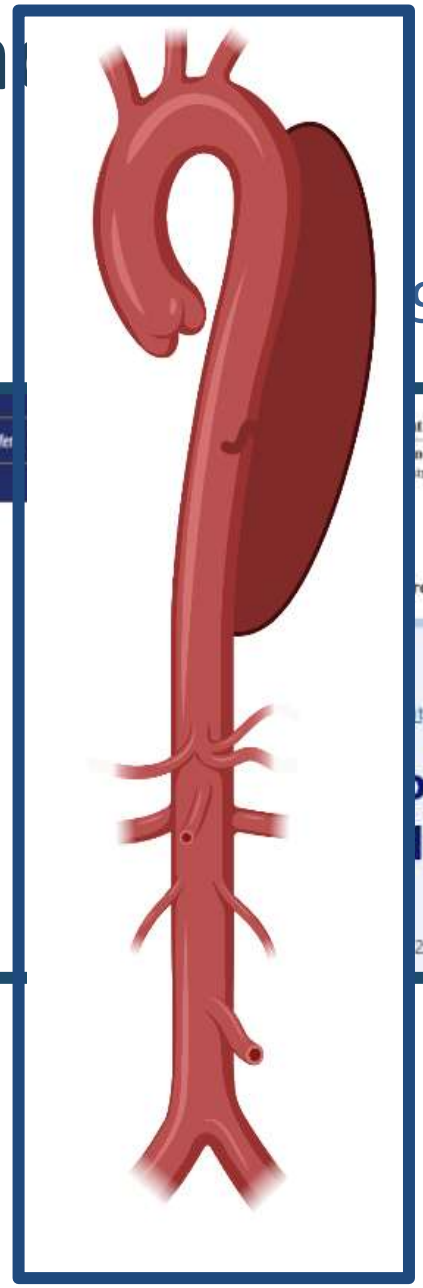


April 2019:

December 2018:

9:





What evidence?

- Safety announcement based on:
  - Cases reported to FDA
  - 4 published studied providing moderate quality evidence
    - Inconsistencies in study designs (active comparators, follow-up, patient age, inclusion of specific fluoroquinolones)
    - Conflicting results

- Researchers aware of limitations of prior studies

What did we do?

Study

1. **Characterization:** incidence and time-to event of aortic events following fluoroquinolone exposure
2. **Estimation:** comparative safety of fluoroquinolones vs other antibiotics

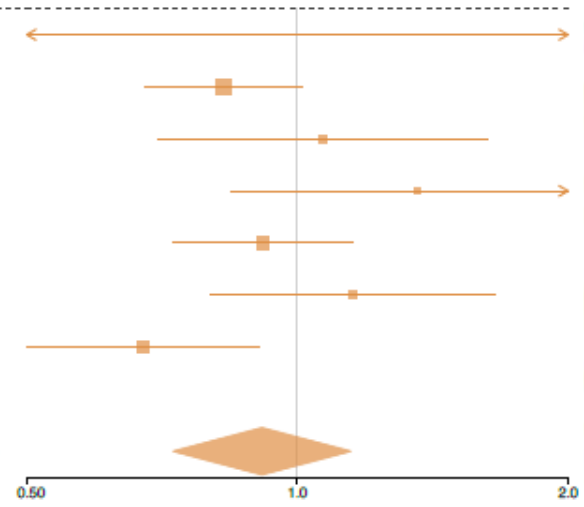
Were people **more likely to go to hospital for internal bleeding within 60 days** of starting certain antibiotic treatment **for treatment of urinary tract infections** compared to people who received other commonly used antibiotics?

Findings?

- Implemented OHDSI's objective study diagnostics
- > 13.5 million people
- 14 different databases from 5 different countries

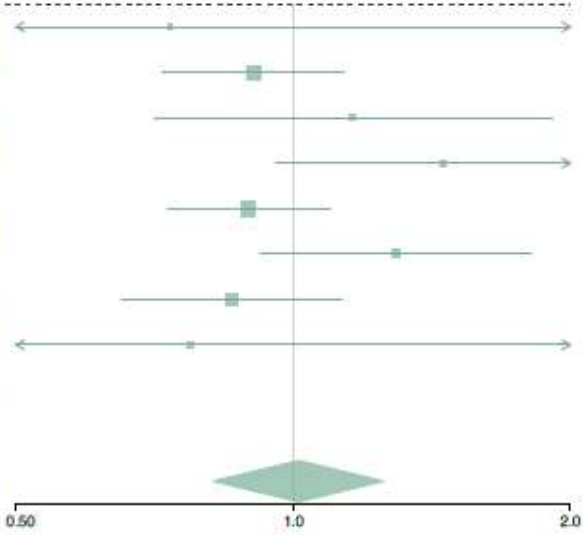
FQ vs TMP

Source	Event Rate FQ	Event Rate TMP	Hazard ratio (95% CI)
CUIMC(US)	<7.69	7.69	0.45(0.08 to 2.58)
Clinformatics(US)	3.75	4.53	0.83(0.68 to 1.01)
IBM CCAE(US)	0.58	0.53	1.07(0.70 to 1.63)
IBM MDCD(US)	2.87	2.09	1.36(0.84 to 2.19)
Optum EHR(US)	2.23	2.34	0.92(0.73 to 1.15)
PharMetrics(US)	1.12	0.96	1.15(0.80 to 1.66)
VA(US)	4.46	6.40	0.67(0.50 to 0.91)
Summary	**<2.09**	**2.31**	**0.91(0.73 to 1.15)**



FQ vs CPH

Source	Event Rate FQ	Event Rate CPH	Hazard ratio (95% CI)
CUIMC(US)	6.31	8.84	0.73(0.22 to 2.42)
Clinformatics(US)	6.29	7.03	0.90(0.72 to 1.13)
IBM CCAE(US)	1.01	0.85	1.16(0.71 to 1.91)
IBM MDCD(US)	5.19	3.52	1.45(0.96 to 2.21)
Optum EHR(US)	4.10	4.50	0.89(0.73 to 1.09)
PharMetrics(US)	2.41	1.84	1.29(0.92 to 1.81)
VA(US)	9.25	10.46	0.86(0.65 to 1.13)
TMUCRD(TW)	<4.08	4.05	0.77(0.19 to 3.08)
Japan Claims(JP)	<2.06	0.00	NA
Summary	**<4.07**	**4.18**	**1.01(0.82 to 1.25)**



Findings?

SAFE

SAFE

SAFE

Mahidol University
Faculty of Medicine
Siriraj Hospital

Siriraj Informatics and
Data Innovation Center
SiData

Mapping Thai Medicine Terminology to RxNorm: Lesson Learned in Standard Vocabulary Integration

Krittaphas Chaisutyakorn (Krit)
Thailand

Introduction

- Siriraj Hospital is currently transforming its EHR data to OMOP CDM
- Current vocabulary
  a. Luckily, we also have Thai Medicine Terminology (TMT) codes which is maintained and use in nation-wide level
  b. But we still didn't have international/standard level codes
- TASK: Mapping Thai Medicine Terminology (TMT) to RxNorm codes

Breaking down the task

1. At which level

2. What codes to map

3. How to map

At which levels?

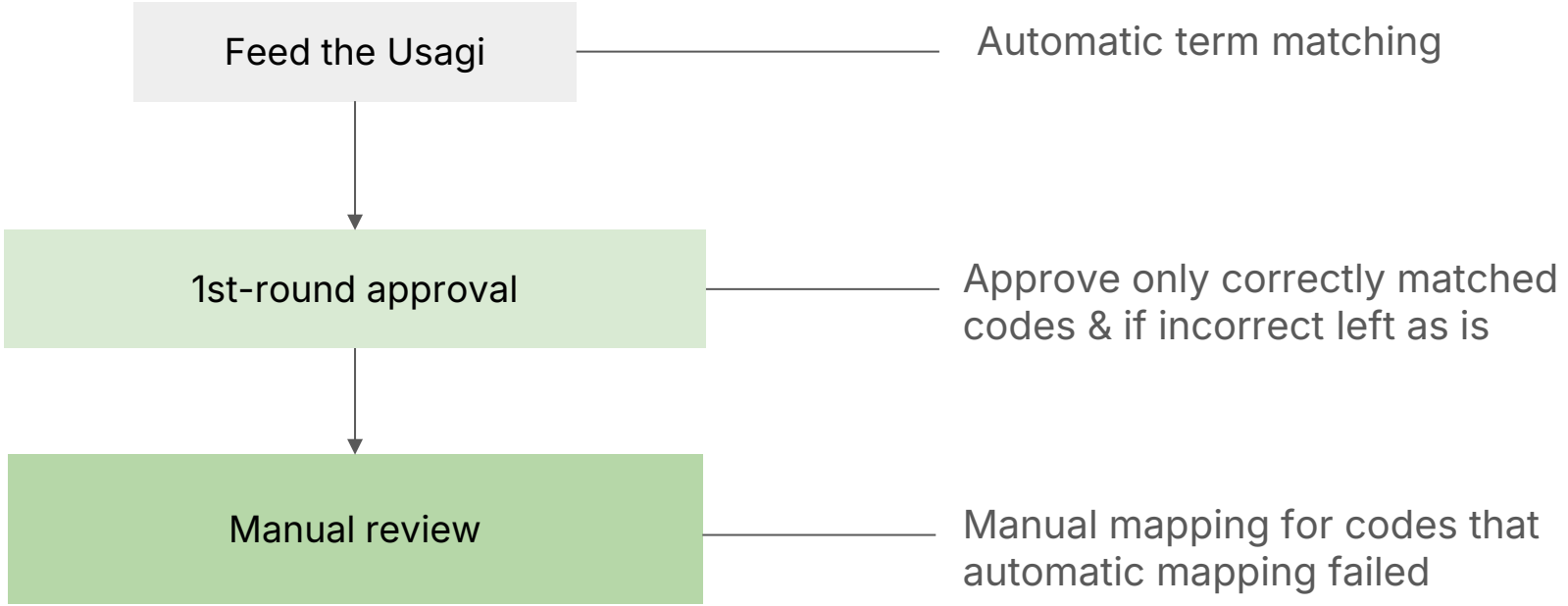
- Codes which contain Ingredient + Strength + Form are sufficient for most researches

TMT	RxNorm
amoxicillin 875 mg + clavulanic acid 125 mg film-coated tablet, 1 tablet	amoxicillin 875 MG / clavulanate 125 MG Oral Tablet

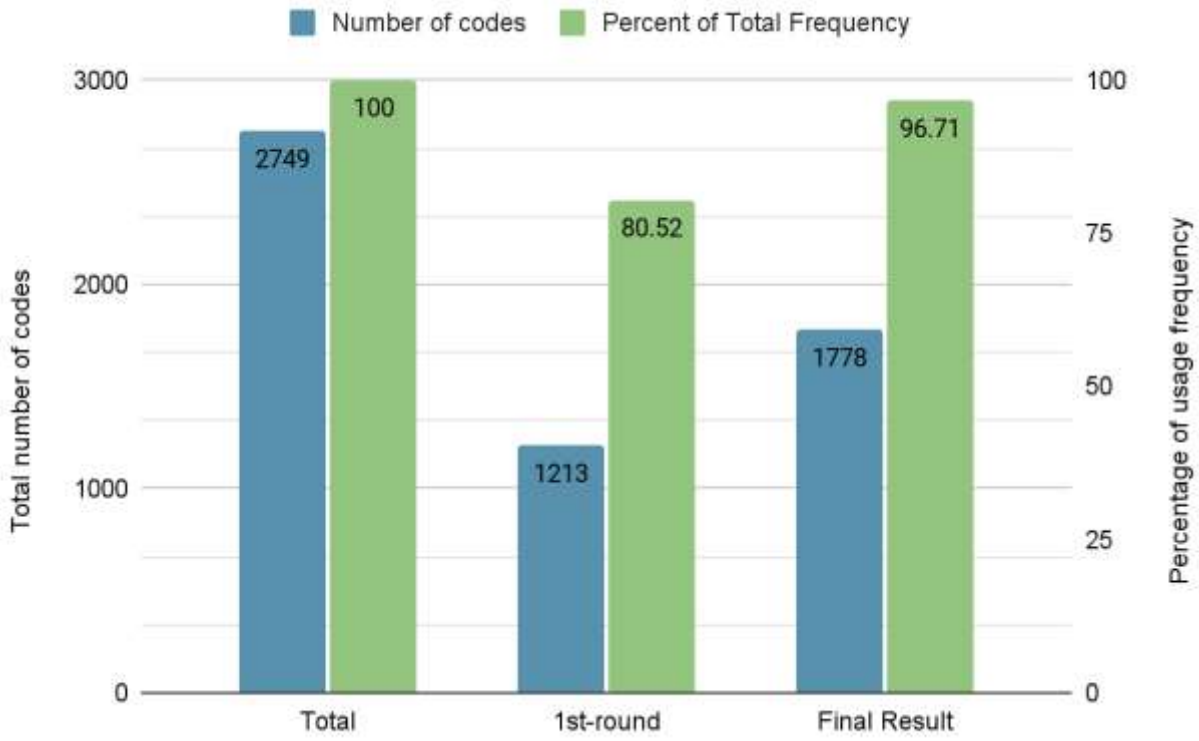
What codes to map

- Siriraj Hospital have 2,749 unique medication codes
- We aim to cover 95% of medication usage frequency

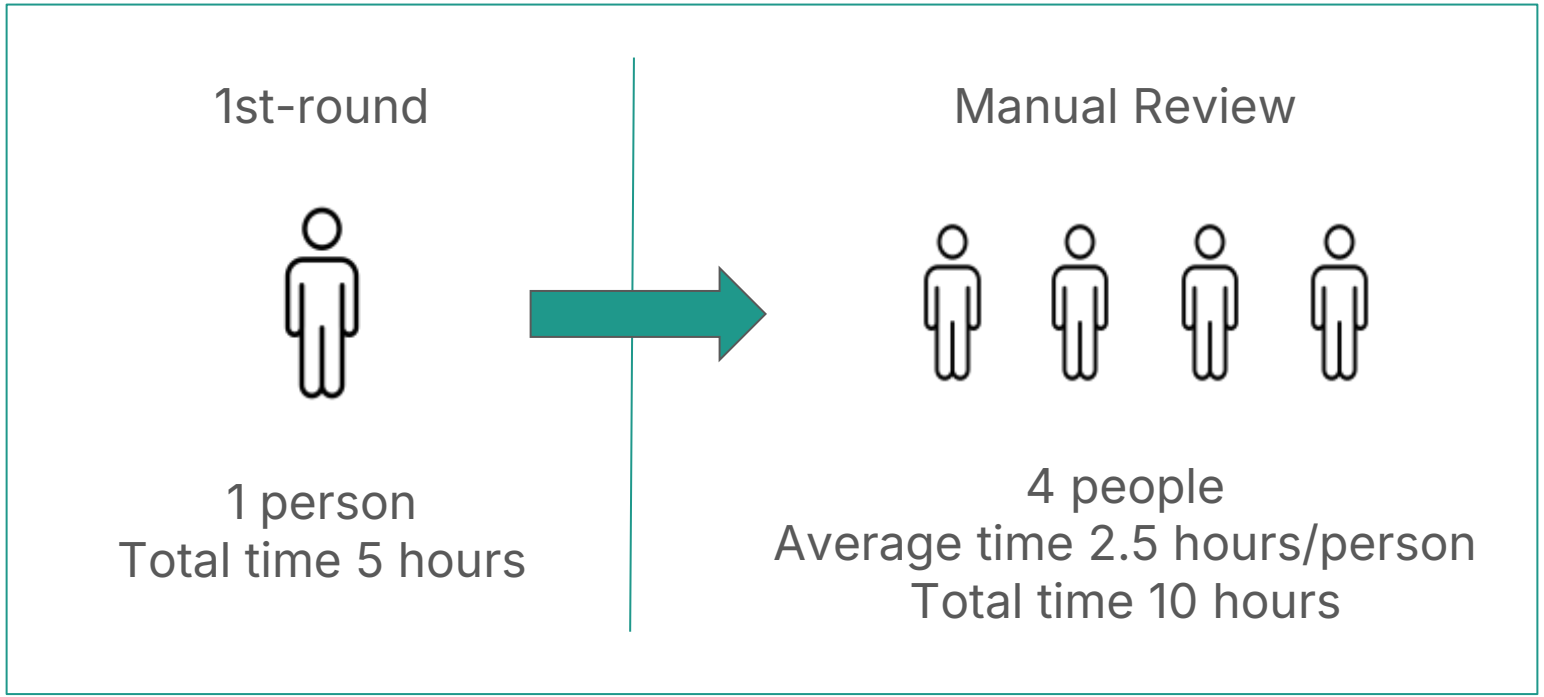
How to map


Feed the Usagi
Automatic term matching
1st-round approval
Approve only correctly matched codes & if incorrect left as is
Manual review
Manual mapping for codes that automatic mapping failed

Number of codes and Percentage



Time investment



All of this processes was done in 3 weeks

Lesson Learned

- Challenging medications to be mapped are intravenous medications, combined medications, vaccines
- 60% of medications contribute to 96% of total usage
- Seemingly impossible task can be broken down to smaller achievable tasks, making the whole process much more feasible.

Thank You

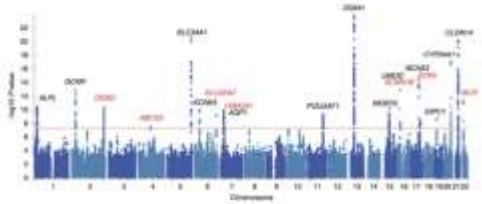
contact: krittaphas.cha@mahidol.edu

Enabling Genomic Data Harmonization in OMOP CDM

Erwin Tantoso[1], Kee Yuan Ngiam[2,3], Mukkesh Kumar[1,4]

[1] Bioinformatics Institute, Agency for Science Technology and Research, Singapore
[2] Division of General Surgery (Thyroid & Endocrine Surgery), National University Hospital Singapore, Singapore
[3] Department of Biomedical Informatics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore
[4] Institute for Human Development and Potential, Agency for Science Technology and Research, Singapore

OMOP CDM enable standardized analytics across a distributed data network



Discovery-based Analysis

Targeted Gene Panel

List of significant variants

Disease Cohort Data

ETL

OHDSI Data Network

Site 1

ETL

CDM 2

Site 2

ETL

CDM 3

CDM Local

OHDSI Open-Source Analytic Codes

OHDSI-Atlas

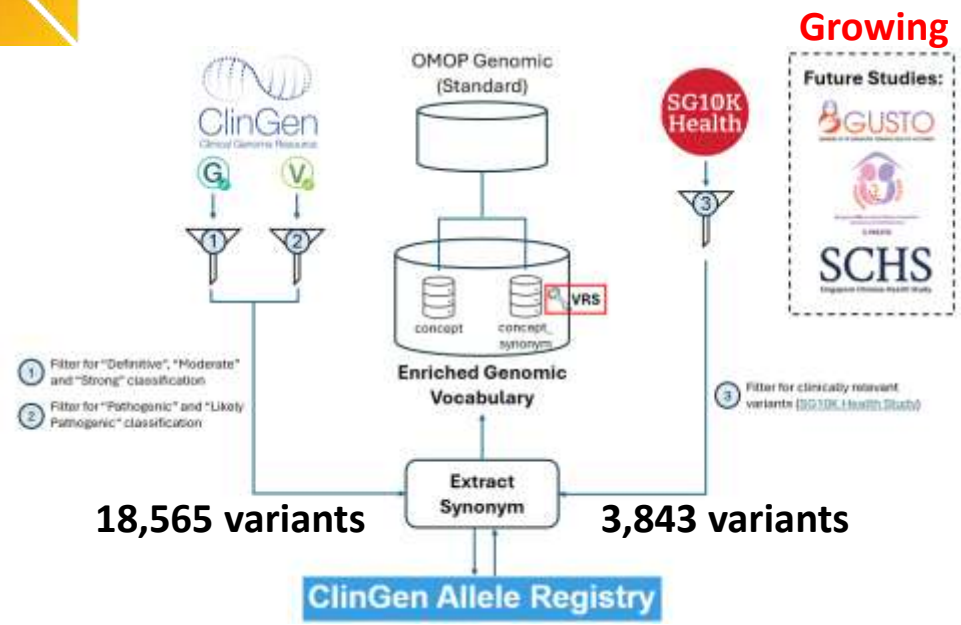
Analysis Results (Open Evidence)

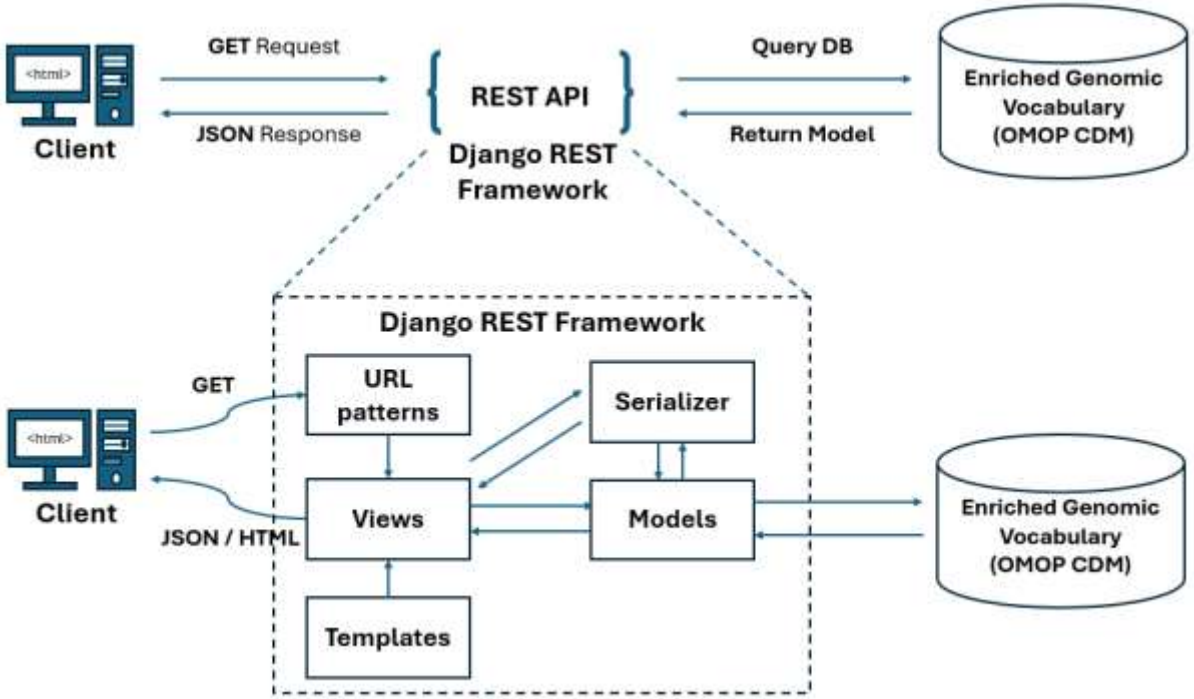
OMOP Genomic for Genetic Data Harmonization in OMOP CDM

- **Challenges:**
  - Limited coverage:
    - OMOP Genomic is designed for oncology research (Golozar A. and Reich C. 2022)
    - Limited coverage on clinically relevant variants in Asian population, especially Singapore
  - Technical challenges:
    - Mapping tool for diverse variant representation
    - OMOP CDM conversion

- **Our objectives:**
  1. Enriching the genomic vocabulary with clinically relevant variants from publicly available literature/curated datasets in the local Singapore context
  2. Allow streamlined transformation of variant data to OMOP CDM space
     a) Query and mapping tool from variant to OMOP Concept IDs
     b) Streamlined annotation of variant data with OMOP Concept IDs
     c) OMOP CDM Genomic Data Transformer

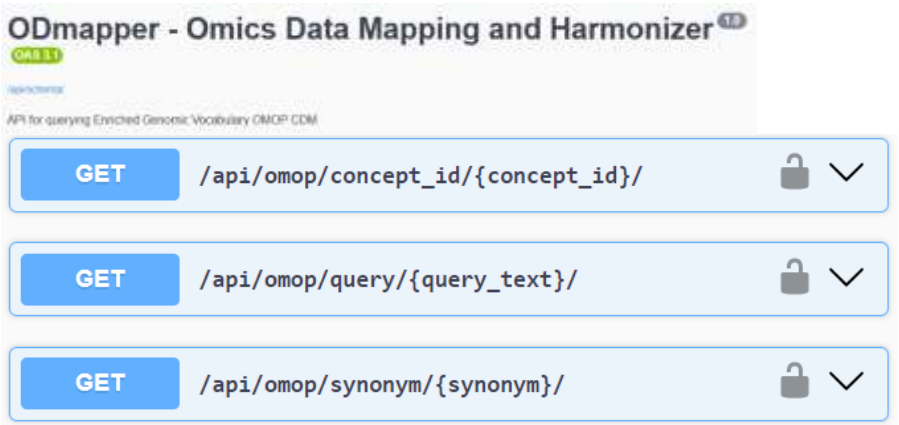
Methods and Results





Genes related to phenotype (ACMG v3.2)	#Genes	OMOP Genomic	Enriched Genomic Vocabulary
Cancer phenotype	28	28	28
Cardiovascular phenotypes	40	10	40
Inborn errors of metabolism phenotypes	4	1	4
Miscellaneous phenotypes	9	7	9

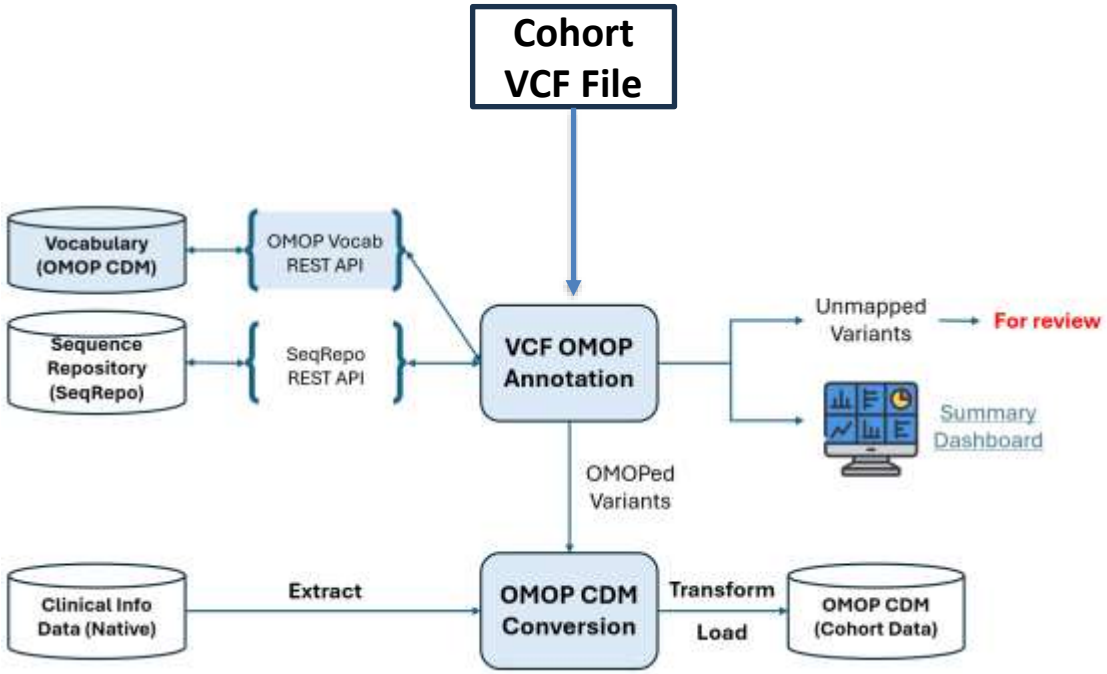
Enriched Genomic vocabulary contributes towards a more comprehensive coverage list



https://github.com/biierwint/ODmapper

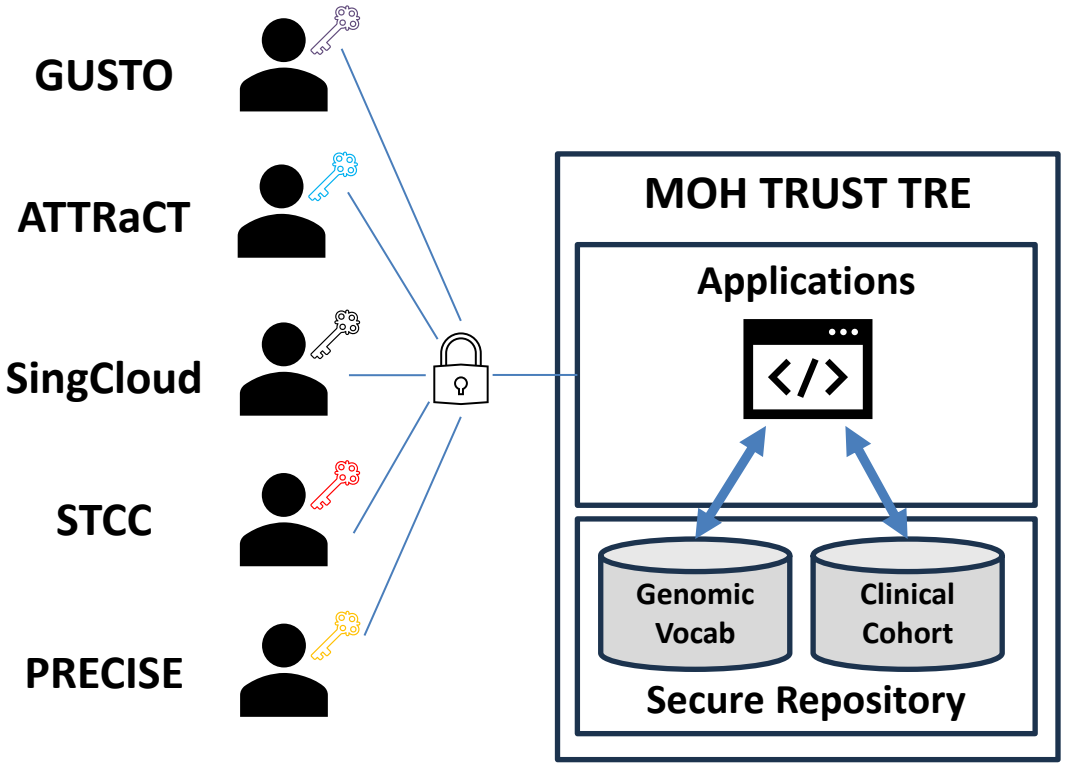
Future Work

First: Development of automated OMOP CDM converter for genomic data



The cohort VCF file will be annotated with OMOP concept_id based on enriched genomic vocabulary. The OMOPed variants will be converted to OMOP CDM (v5.4)

Second: Deployment of application on the MOH-TRUST TRE (enTRUST)



MOH TRUST TRE as the central TRE to host the OMOP CDM enriched genomic vocabulary and sensitive clinical cohorts which can only be accessible by the trusted users.

Acknowledgement

- **NUHS and MOH TRUST**
  - Prof. Ngiam Kee Yuan
  - Ms. Koh Mingshi
  - Dr. Low Pin Yan
  - Mr. Yeo Zhenxuan

- **GUSTO Team**
  - Dr. P Mukkesh Kumar
  - Ms. Cindy Ho
  - Mr. Tan Jing Yang

- **PRECISE Team**
  - Dr. Nicolas Bertin
  - Dr. Maxime Hebrard
  - Dr. Max Lam

- **OHDSI Team**
  - Dr. Anna Ostropolets
  - Dr. Christian Reich
  - Mr. Laurence Lawrence-Archer (Odysseus Inc.)

Asian and/or Pacific Islander: Unmasking health disparities within commonly aggregated diverse populations in the US Department of Veterans Affairs (VA)

Benjamin Viernes[1,3], Scott L DuVall[1,2], Patrick R Alba[1,2], Qiwei Gan[1,2], Elizabeth E Hanchrow[1], Mengke Hu[1,2], Gregorio Coronado[1,3], Andy Subica, Curtis Lowery[3], Scott Hofer[3,4], Vicki Shambaugh[3,4], Kalani Raphael[1-3]

1 – VA Salt Lake City Health Care System, Salt Lake City, UT, USA

2 – Department of Internal Medicine, University of Utah Medical School, Salt Lake City, UT, USA

3 – Center for Pacific Islander Veterans Health. VA Pacific Islands Healthcare System. Honolulu, HI, USA

4 – Pacific Health Research and Education Institute. Honolulu, HI, USA

OHDSI
OBSERVATIONAL HEALTH DATA SCIENCES AND INFORMATICS

VA Research
DISCOVERY ★ INNOVATION ★ ADVANCEMENT
VA Informatics and Computing Infrastructure

THE UNIVERSITY OF UTAH®

VA | U.S. Department of Veterans Affairs

United States' Government Office of Management and Budget (OMB) Race Categories

- 1977 Directive 15: Asian or Pacific Islander (API)
- 1997 Statistical Policy Directive 15: First step towards disaggregating Asian or Pacific Islander to 2 distinct categories
  - Asian
  - Native Hawaiian or Pacific Islander (NHPI)
- 2024: API still used in many common data sources

Failure to Recognize Pacific Islander Inequity

- Of 853,654 VA Patients identified as API, 180,804 can be identified by the more granular NHPI
- NHPI that can be disaggregated are different than the aggregate API category.

	API	NHPI
VA Patients	853,654	180,804
Mean Age	52.74	60.91
Sex = Male - N (%)	746,646 (87.5%)	158,663 (87.8%)
Mean Years of VA Observation	12.3	13.6
Mean Measurement Count	224.5	316.5
Mean Diagnosis Count	1097	1614
Mean VA Interaction Count	175	236

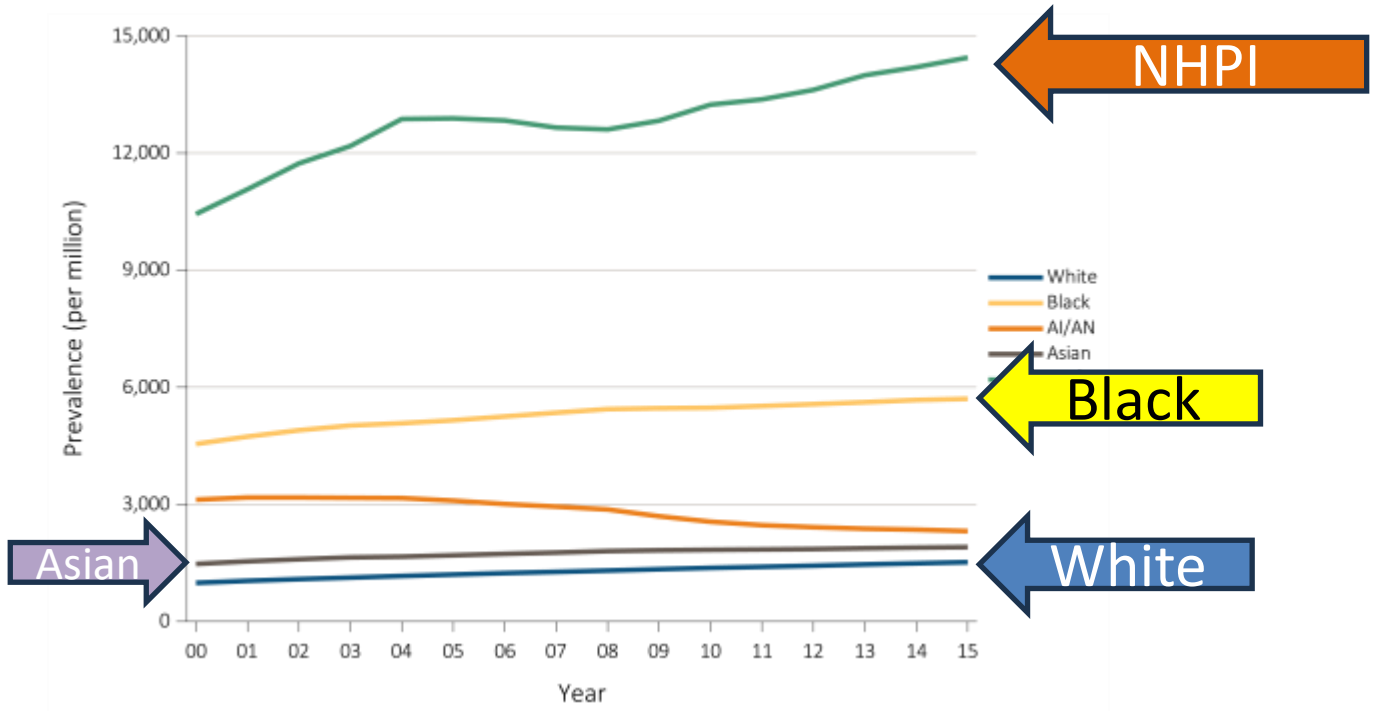
Masking Health Disparities

- Disaggregated from API, NHPI exhibit higher proportions with cardio-metabolic diseases
- Disaggregation of NHPI demonstrates the need to enforce disaggregated collection of race and ethnicity data

	API	NHPI			
			Polynesian	Micronesian	Melanesian
VA Patients	853,654	180,804	48,890	17,864	2,383
NHPI-related Conditions - N (%)					
Chronic Kidney Disease	106,797 (12.5%)	**43,186 (23.9%)**	**14,647 (30.0%)**	4,277 (23.9%)	263 (11.0%)
Cardiovascular Disease	118,816 (13.9%)	**51,887 (28.7%)**	**14,635 (29.9%)**	4,139 (23.2%)	269 (11.3%)
Diabetes Mellitus	185,285 (21.7%)	**66,285 (36.7%)**	**21,092 (43.1%)**	6,282 (35.2%)	440 (18.5%)

NHPI in 2017 US Renal Data Systems Report

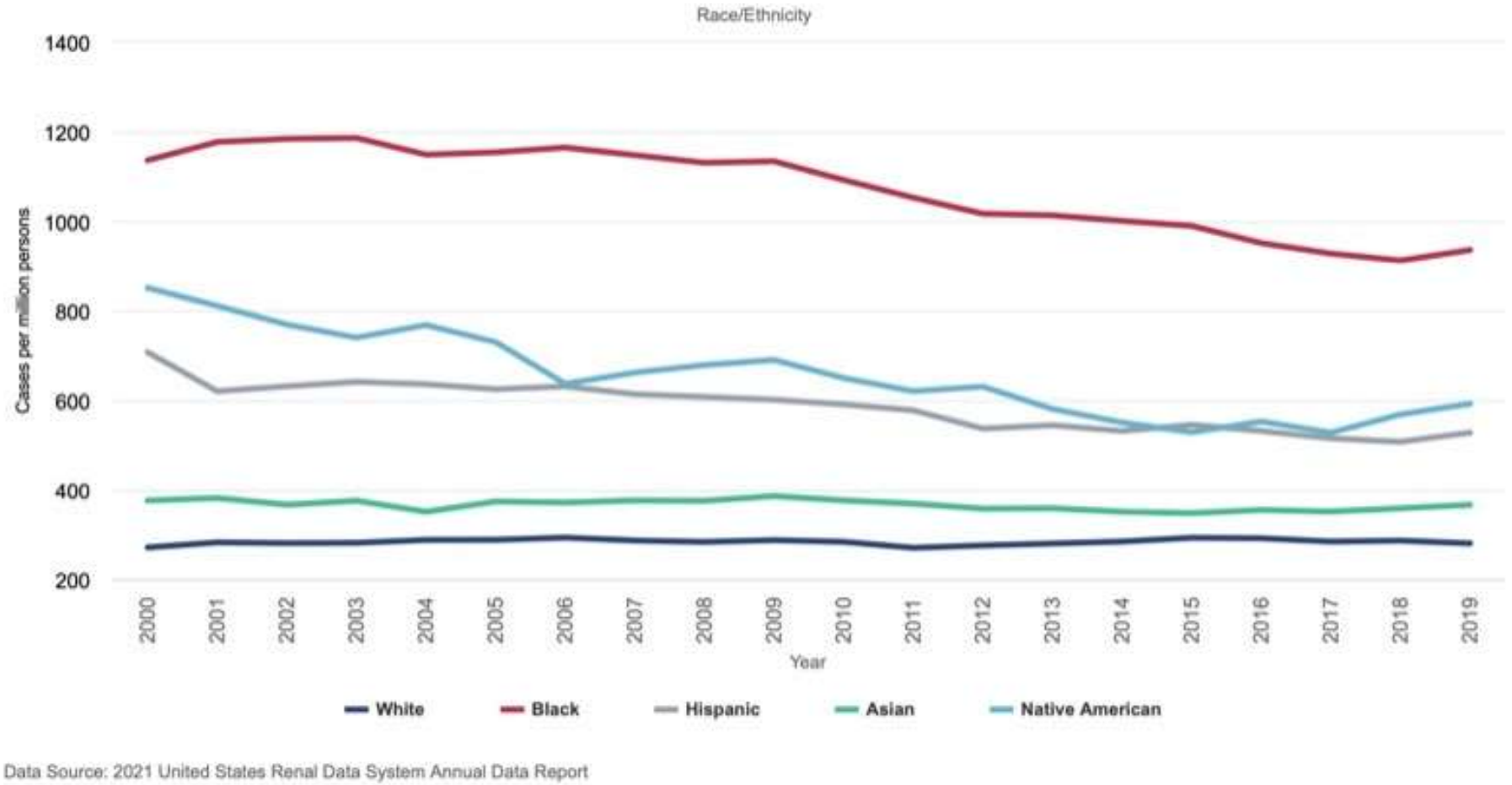
- Prevalence of ESRD in 2017



Data Source: Reference Table B.2(2) and special analyses, USRDS ESRD Database. Point prevalence on December 31 of each year. Standardized for age and sex. The standard population was the U.S. population in 2011. Abbreviations NH/PI: Native Hawaiian/Pacific Islander; AI/AN: Americans Indian/Alaska Natives; ESRD, end-stage renal disease.

NHPI invisible in 2021

- NHPI absent from USRDS reports of Kidney failure in the US by race & ethnicity



Data Source: 2021 United States Renal Data System Annual Data Report

Applying the OMOP Common Data Model to Facilitate Benefit-Risk Assessments of Medicinal Products Using Real-World Data from Singapore and South Korea

Hui Xing Tan, Desmond Chun Hwee Teo, Dongyun Lee, Chungsoo Kim, Jing Wei Neo, Cynthia Sung, Haroun Chahed, Pei San Ang, Doreen Su Yin Tan, Rae Woong Park, Sreemanee Raaj Dorajoo

Vigilance & Compliance Branch, Health Products Regulation Group, Health Sciences Authority, Singapore
Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Korea
Department of Biomedical Sciences, Graduate School of Medicine, Ajou University, Suwon, Korea
Health Services and Systems Research, Duke-NUS Medical School, Singapore
Department of Pharmacy, Khoo Teck Puat Hospital, Singapore

Agenda

- Background
- Objectives
- Methods
- Results
- Discussion
- Conclusion

Background

- Regulators need to monitor the effectiveness, safety, and utilization of health products in routine care
- There is increasing interest among regulators in using real-world data (RWD) to enhance their understanding of the benefits and risks of health products
- However, analysing RWD (typically from healthcare databases) and generating real-world evidence (RWE) to fulfil this can be challenging

Objectives

- Characterise the benefits of converting Electronic Medical Records (EMRs) to a common data model (CDM)
- Assess the potential of CDM-converted data to rapidly generate insights for benefit-risk assessments in post-market regulatory evaluations and decisions

Methods

Two phases:

1. Carry out conversion of Electronic Medical Record (EMR) data from their source files to the OMOP-CDM
2. Simulate an illustrative benefit-risk assessment of the converted data using available tools and code sets

Methods

The data:

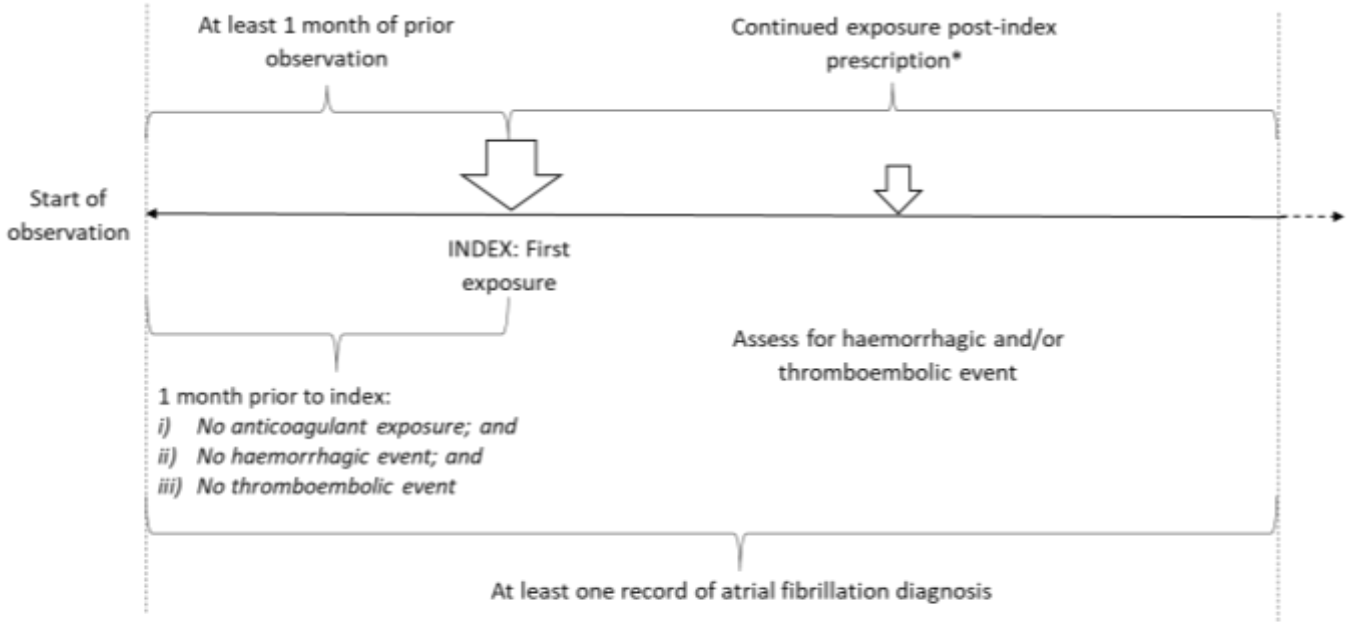
	Singapore	South Korea
Data source	EMR data from a tertiary acute care hospital in Singapore which provides a wide range of medical and surgical specialty services	EMR data from a 1,200-bed tertiary care facility in South Korea providing medical and surgical specialty services
Population size	260,000 unique patients	2,700,000 unique patients
Period of data	Jan 2013 – Dec 2016	Jan 1994 – Dec 2020

Methods

- Conversion of Singapore source data to OMOP-CDM version 5.3.0
  - **Used "Rabbit-In-a-Hat" software**
  - Vocabulary mappings were applied to translate the codes and values used in the source data to those used in the CDM.
    - Diagnosis codes: ICD-9, ICD-10 → SNOMED-CT
    - **Drugs: RxNorm Extension**
    - Laboratory tests and vitals measurements: Logical Observation Identifiers Names and Codes (LOINC)

Methods

- Illustrative Analysis following CDM Conversion
  - Patients with newly diagnosed **atrial fibrillation (AF)** who were newly started on **oral anticoagulants (OAC)**.
  - Patients were followed up for the occurrence of **bleeding (safety)** or **thromboembolic events (effectiveness)** at any time after the first OAC exposure.



At minimum, one prescription for an oral anticoagulant within a 3-month window

Methods

- Illustrative Analysis following CDM Conversion
  - Produced bar charts to visualize drug utilization and effectiveness and safety event proportions to facilitate multiple comparisons in benefit-risk assessments.
  - External validation on EMR data from Ajou University Medical Center (AUMC) in South Korea was performed
    - to illustrate the potential of generating comparable results from different geographical cohorts of patients, and
    - to assess whether any signal of observable differences between agents persisted across different cohorts.

Results

Conversion of Singapore source Data to OMOP-CDM version 5.3.0

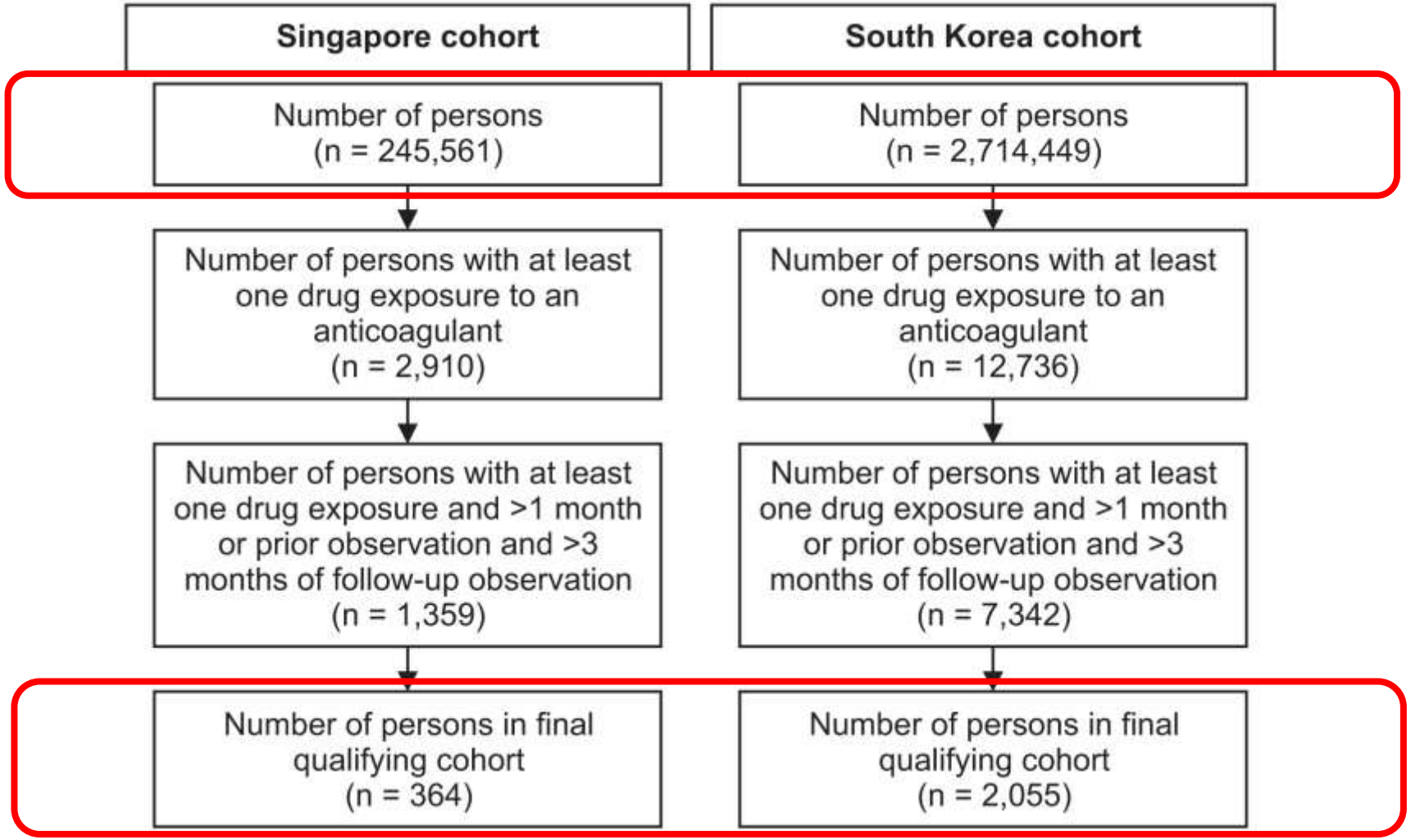
Quantity and structure of data imported from a tertiary acute care hospital in Singapore from January 2013 to December 2016

OMOP-CDM table		Source table		
Table name	Number of rows of records	Table name	Number of rows of records	Proportion migrated (%)
person	245,561	t_demographics	258,038	95.2
condition_occurrence	(primary) 210,830	t_primary_diagnosis	222,554	94.7
	(secondary) 799,169	t_secondary_diagnosis	839,265	95.2
measurement	14,116,544	t_lab_result	15,523,576	90.9
visit_occurrence	1,041,587	t_encounter	1,057,263	98.5
drug_exposure	4,378,657	t_eprescription_dispensing[a]	2,147,505	84.8
		t_inpatient_med_order[b]	3,015,159	84.8

[a]Refers to outpatient pharmacy orders and inpatient discharge prescriptions.

[b]Refers to medications used during inpatient ward stay.

Results

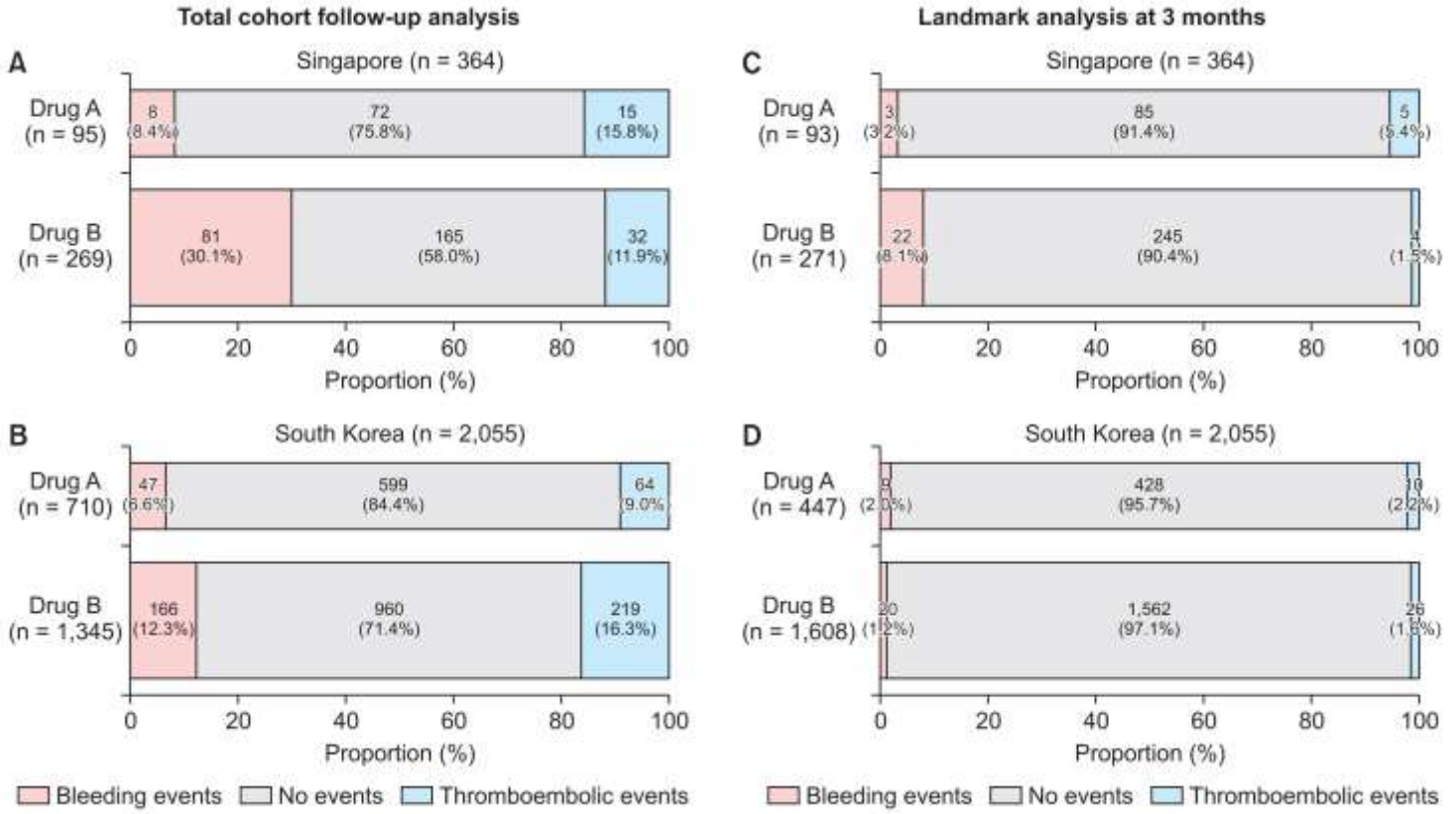

Singapore cohort
South Korea cohort
Number of persons (n = 245,561)
Number of persons (n = 2,714,449)
Number of persons with at least one drug exposure to an anticoagulant (n = 2,910)
Number of persons with at least one drug exposure to an anticoagulant (n = 12,736)
Number of persons with at least one drug exposure and >1 month or prior observation and >3 months of follow-up observation (n = 1,359)
Number of persons with at least one drug exposure and >1 month or prior observation and >3 months of follow-up observation (n = 7,342)
Number of persons in final qualifying cohort (n = 364)
Number of persons in final qualifying cohort (n = 2,055)

Results

Table 2. Baseline characteristics of the final cohorts from Singapore and South Korea

	Warfarin		Rivaroxaban		Combined		p-value[d]
	Singapore	South Korea	Singapore	South Korea	Singapore	South Korea	
Number of patients	269 (73.9)	1,345 (65.5)	95 (26.1)	710 (34.5)	364 (100)	2,055 (100)	
Age (yr)	70 (15)	63 (17)	71 (15)	69 (14)	72 (15)	66 (17)	<0.001
Sex							<0.001
Male	142 (52.7)	854 (63.5)	44 (46.3)	398 (56.1)	186 (51.1)	1,252 (60.9)	
Female	127 (47.2)	491 (36.5)	51 (53.7)	312 (43.9)	178 (48.9)	803 (39.1)	
Race							<0.001
Korean	NA	1,345 (100)	NA	710 (100)	NA	2,055 (100)	
Chinese	163 (60.6)	NA	66 (69.5)	NA	229 (62.9)	NA	
Malay	66 (24.5)	NA	20 (21.1)	NA	86 (23.6)	NA	
Indian	20 (7.4)	NA	5 (5.3)	NA	25 (6.9)	NA	
Others	20 (7.4)	NA	4 (4.2)	NA	24 (6.6)	NA	
Event outcome[a]							<0.001
Bleeding	81 (30.1)	166 (12.3)	8 (8.4)	47 (6.6)	89 (24.5)	213 (10.4)	
Thromboembolic	32 (11.9)	219 (16.3)	15 (15.8)	64 (9.0)	47 (12.9)	283 (13.8)	
Neither	156 (58.0)	960 (71.4)	72 (75.8)	599 (84.4)	228 (62.6)	1,559 (75.9)	

Results



Drug A: rivaroxaban. Drug B: warfarin
Denominator for percentages: number of patients in the cohort on the drug
Numerator for percentages: number of patients in the cohort on the drug who experienced the event (pink/blue), or were event free (grey)

- Initially, rivaroxaban appears to be the preferred agent (left) with fewer bleeding and thromboembolic events
- However, the benefit-risk ratios are similar after adjusting for observation time (right)

Discussion

- CDM conversion alters only the form, but not the substance of the data
  - need to understand the provenance and processes that generated the data
- Conversion can speed up analyses
- Modifications and extensions to previously written code likely required for specific use cases
- Proposed bar graphs remain an unadjusted descriptive analysis of the rate of events in different populations exposed to comparator agents
- Incorporating methods to adjust for confounders and visualize the adjusted event rates would be important areas of future research

Conclusion

- While the structure of the OMOP-CDM and its accessory tools facilitate real-world data analysis, extending them to fulfil regulatory analytic purposes in the post-market setting, such as benefit-risk assessments, may require layering on additional analytic tools and visualisation techniques

Follow HSA on X to receive the latest safety and regulatory updates
@HSAsg

Explore the opinions and attitudes of the application of common data models in regional databases from the perspective of the Chinese population

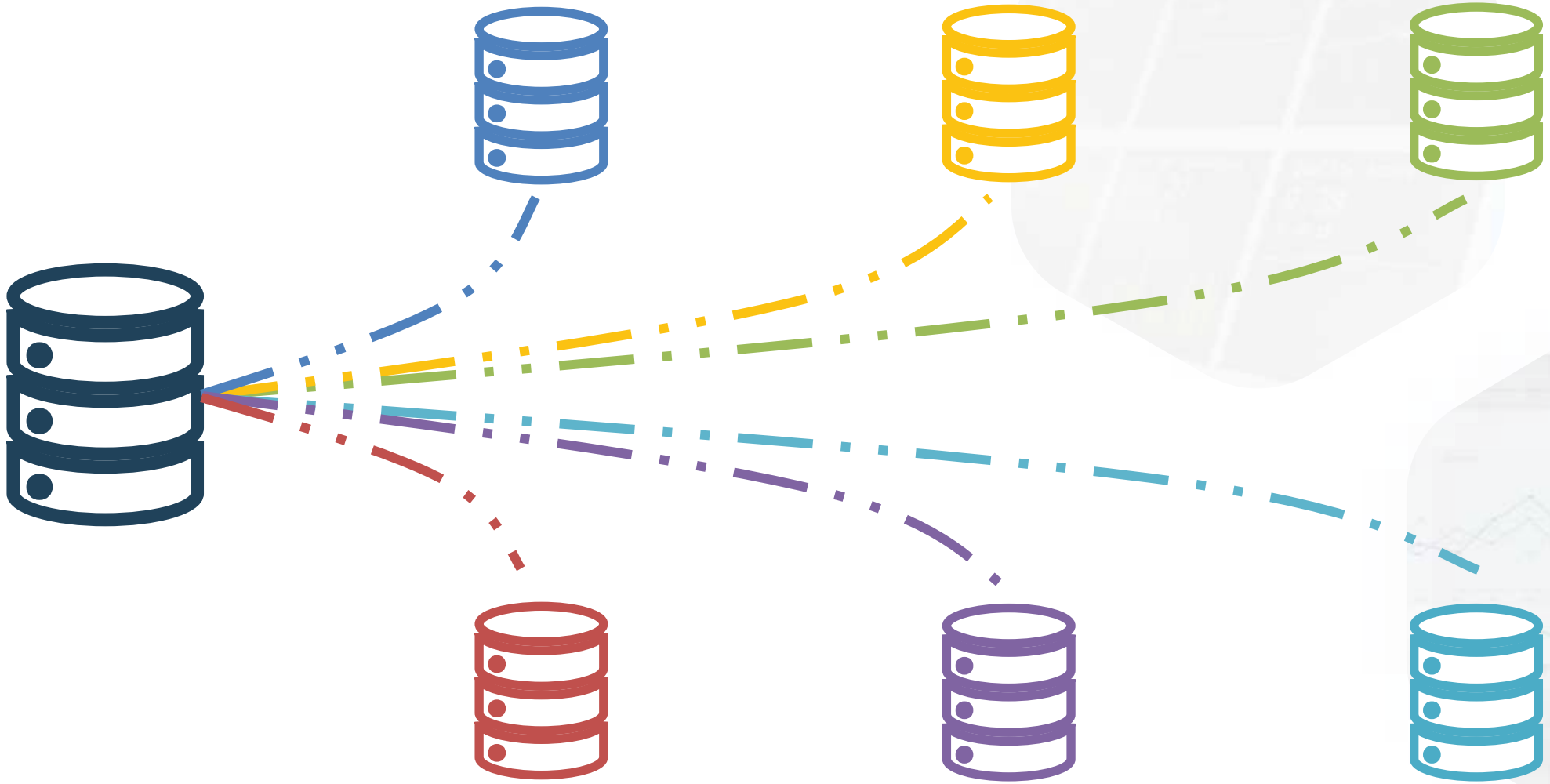
Yexian Yu

Hainan University, China

Hainan Lecheng Institute of Real World Research, China

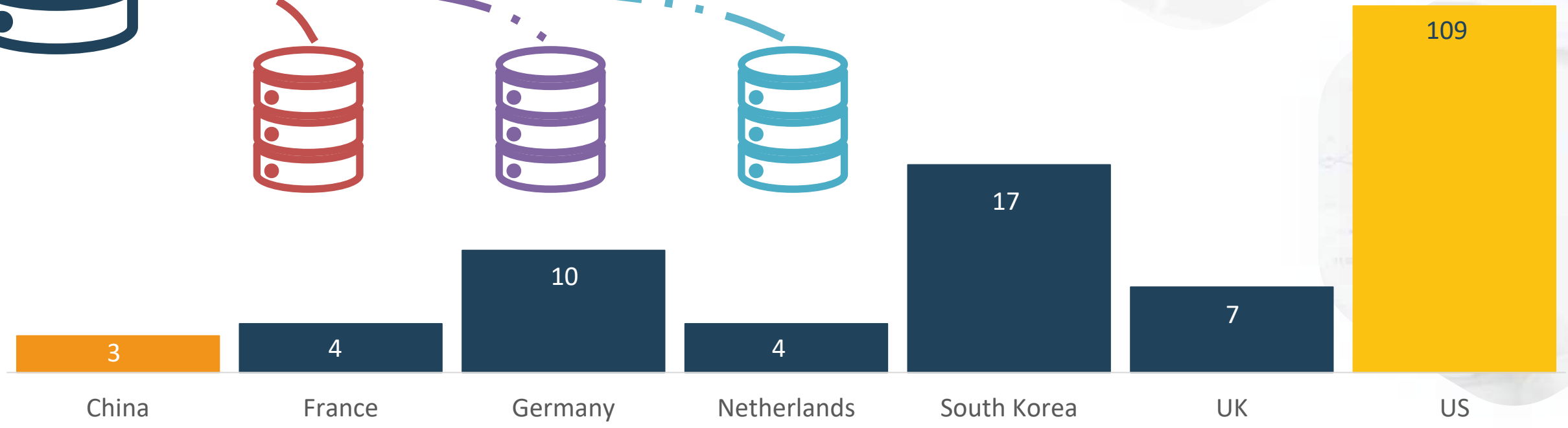
Department of Epidemiology and Biostatistics, School of Public Health, Peking University, China

Background



Background

- The application of CDM in China is **relatively low**, which may be related to a lack of understanding and recognition of CDM[1].



1. Reinecke I, Zoch M, Reich C, Sedlmayr M, Bathelt F. The Usage of OHDSI OMOP - A Scoping Review. Stud Health Technol Inform. 2021 Sep 21;283:9

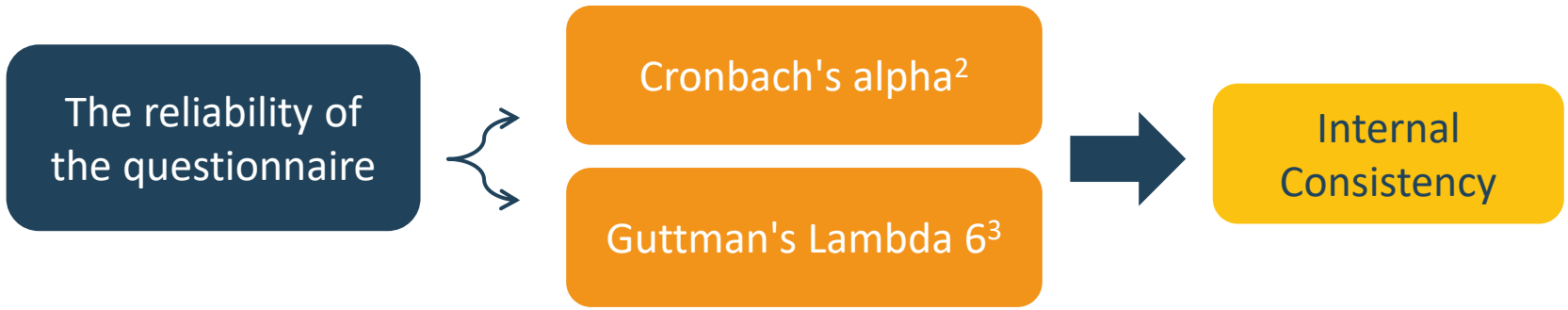
Background

WHY IS THAT?!

Questionnaire Design

Questionnaire Items

Gender	CDM Awareness	OMOP Awareness Source
Age	CDM Awareness Source	OMOP Mainstream Ability
Working/studying Area	CDM Necessity	OMOP Benefits
Occupation	CDM Type	OMOP Challenges
Education Level	OMOP Awareness	Suggestions



2. Cronbach, Lee J. Coefficient alpha and the internal structure of tests. Psychometrika. 1951 Sep;16(3):297–334.
3. Guttman, L. A basis for analyzing test-retest reliability. Psychometrika. 1945 Dec;10(4):255–282.

Survey results

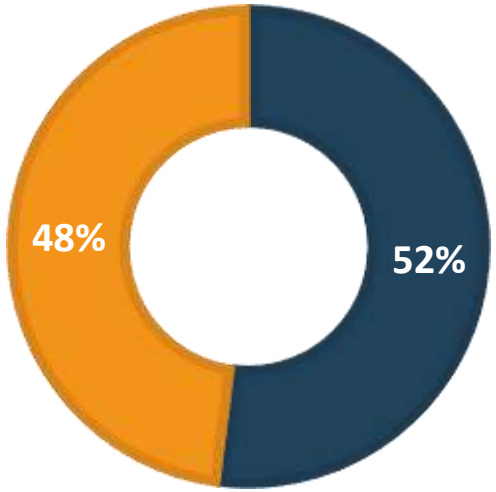
200 Responses

From 18th September 2024 to 22nd September 2024

Cronbach's alpha	0.96
Guttman's Lambda 6	0.98

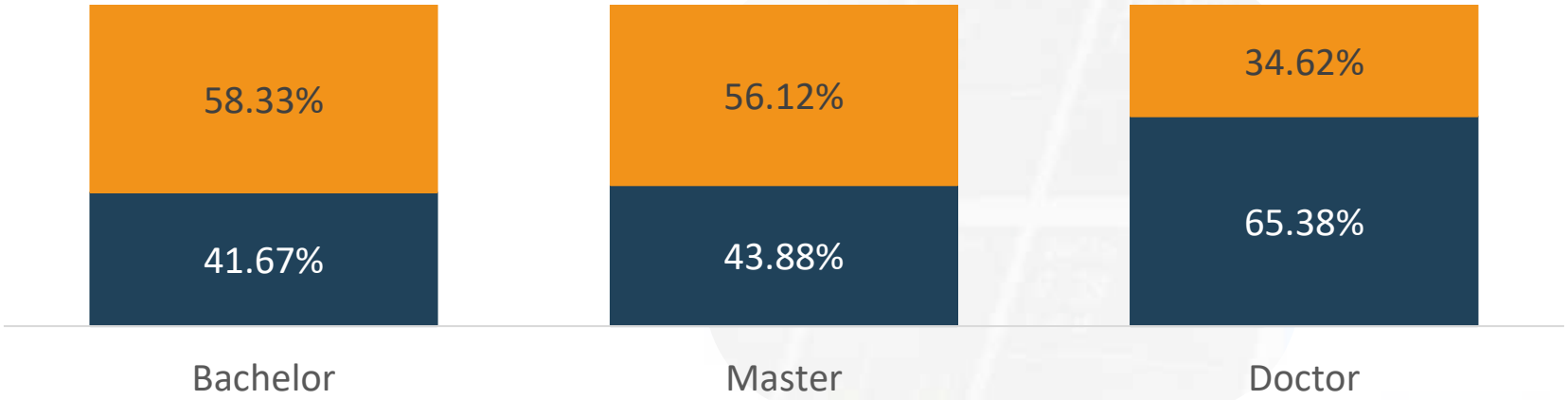
high internal consistency and strong inter-item correlations within the questionnaire

Survey results

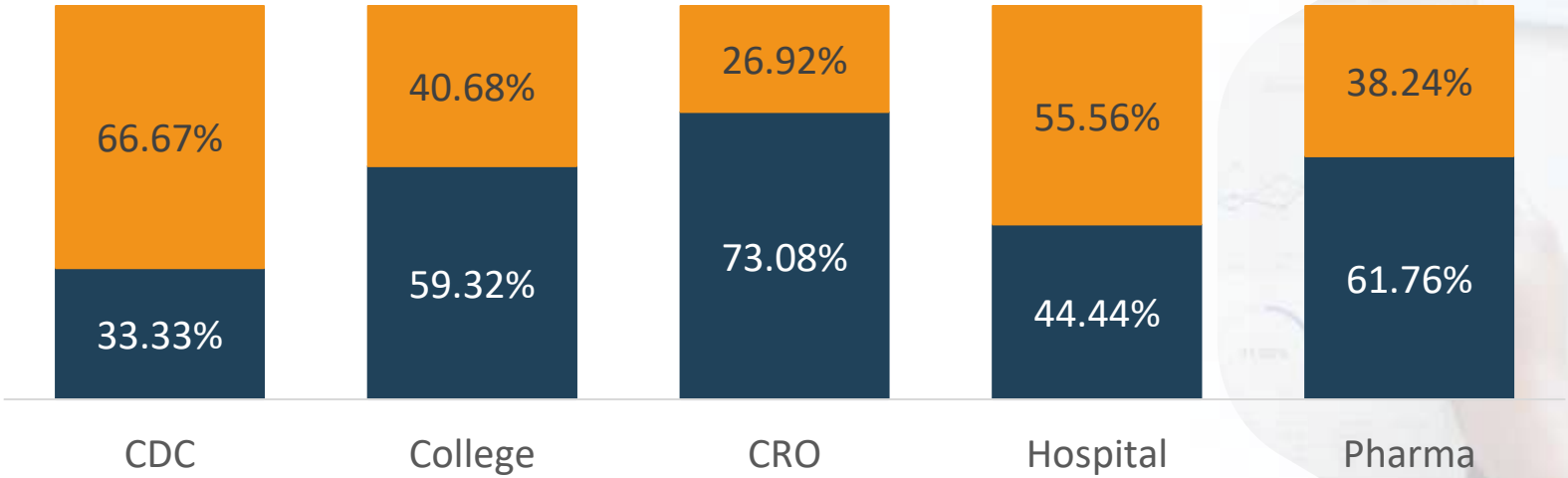


Had not heard of CDM before

Heard of CDM before



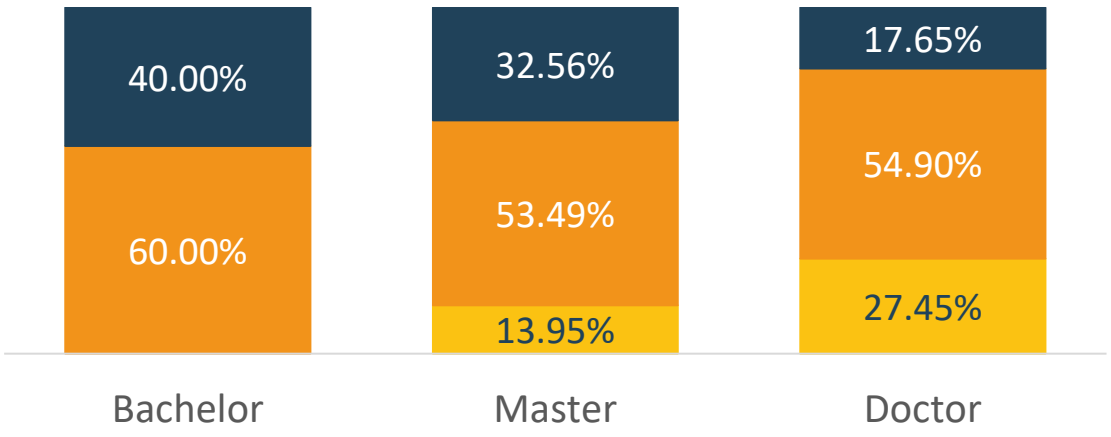
$$\chi^2 = 8.320, P = 0.004$$

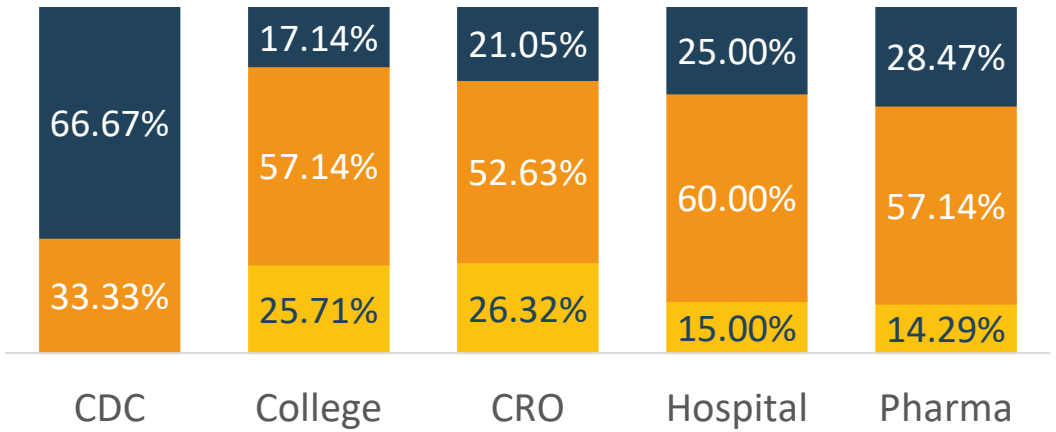


$$\chi^2 = 9.950, P = 0.041$$

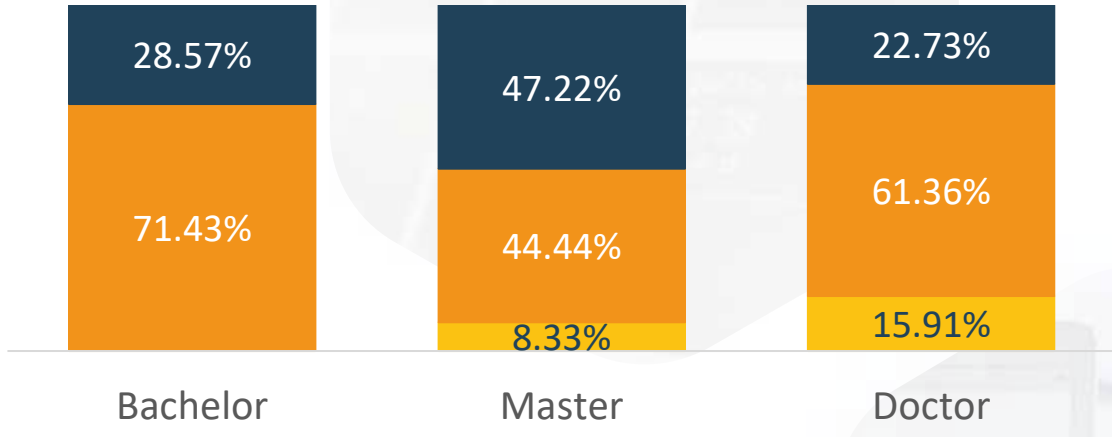
Survey results

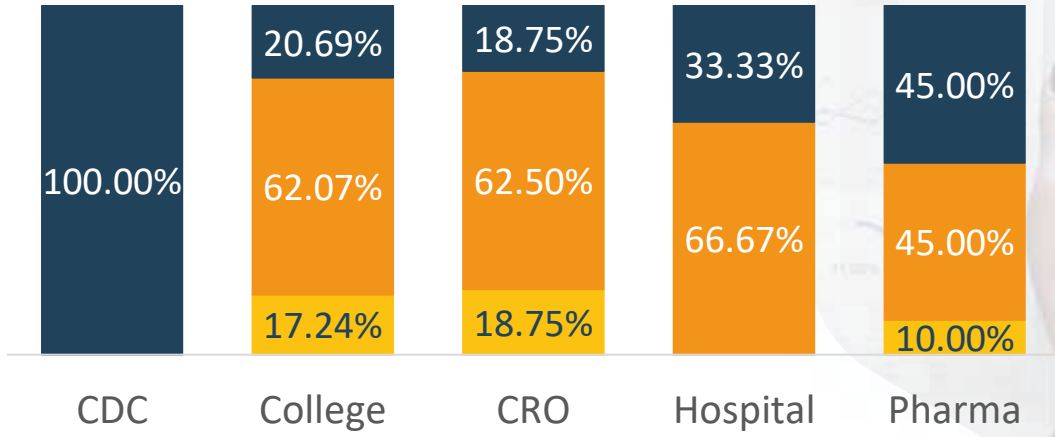
CDM





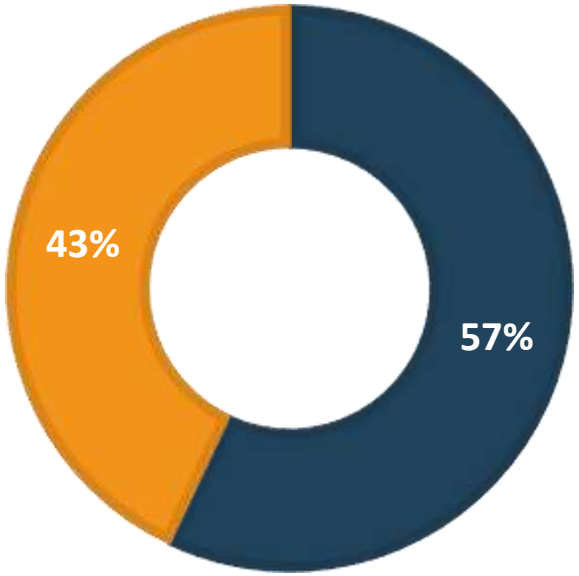
OMOP





Unfamiliar Fairly Familiar Very Familiar

Survey results



- **More than half** of the participants believe that OMOP could become the mainstream choice for CDM in Chinese regional databases.

benefits

- Enhance the operability and comparability of data in Chinese regional databases
- Facilitate the sharing, comparison, and integration of data from different sources.
- Help drive innovation and progress in the research field

Challenges

- There are barriers between different data sources
- Mapping requires a significant amount of manual work, time, and cost
- Some China-specific information cannot be matched to standard concept IDs

Participants' suggestions

Establishing a reasonable data-sharing mechanism at the national level

Establishing small-scale OMOP disease-specific databases

Involving professionals from diverse backgrounds

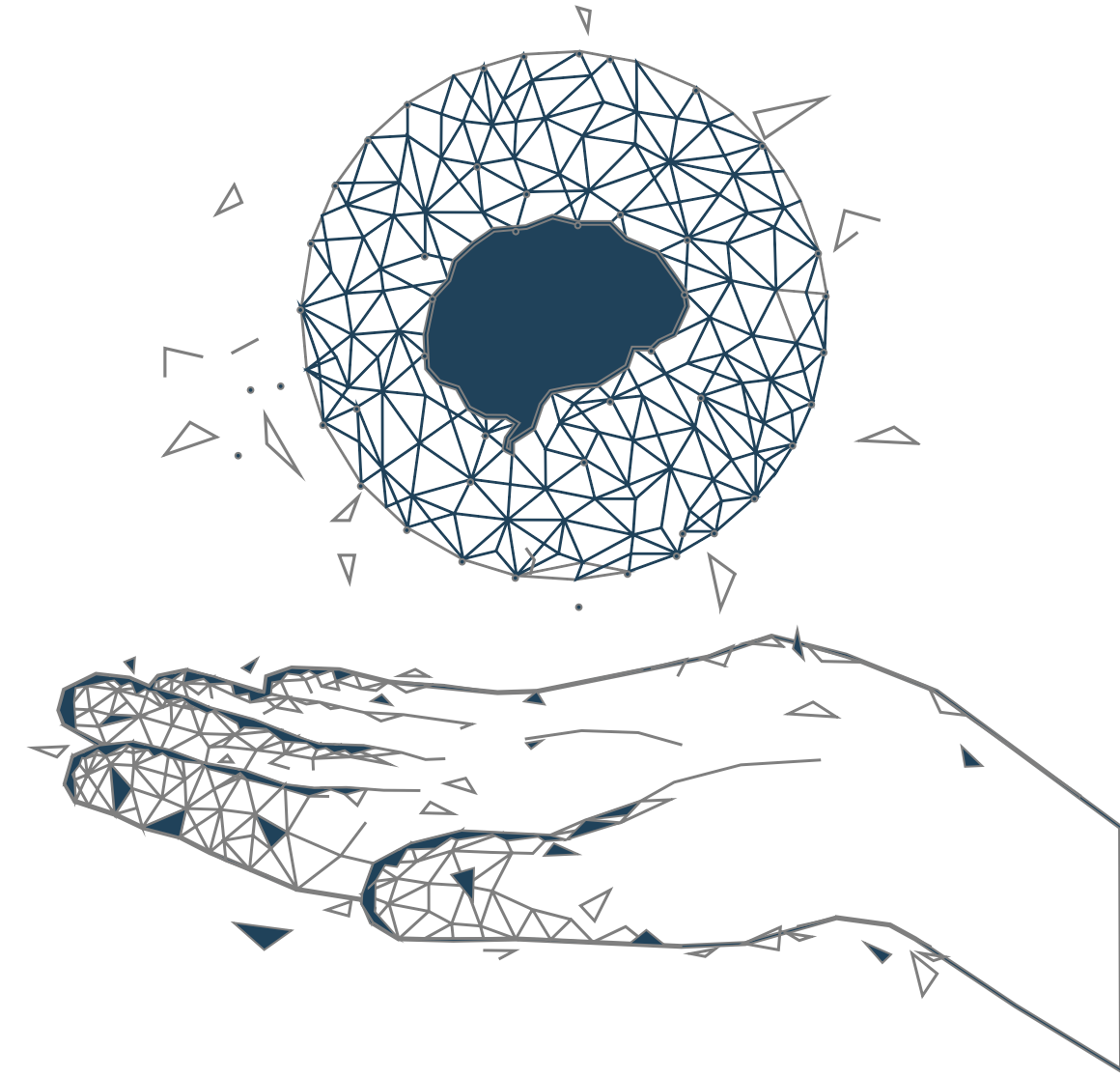
Introduction of artificial Intelligence

Enhancing education and promotion of CDM

Ensuring information encryption and data security

Conclusions

Based on the statistical analysis of the questionnaire results, participants generally hold a **positive attitude** towards the application of CDM in regional databases in China.



Thanks

Yexian Yu: yu.yexian@hainanu.edu.cn