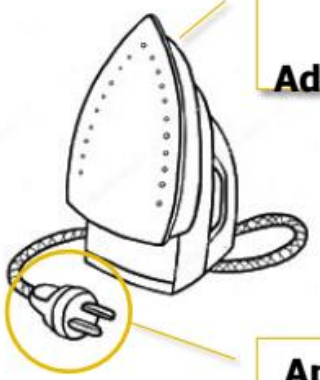
Principles of Phenotyping using OMOP-CDM

Seng Chan You

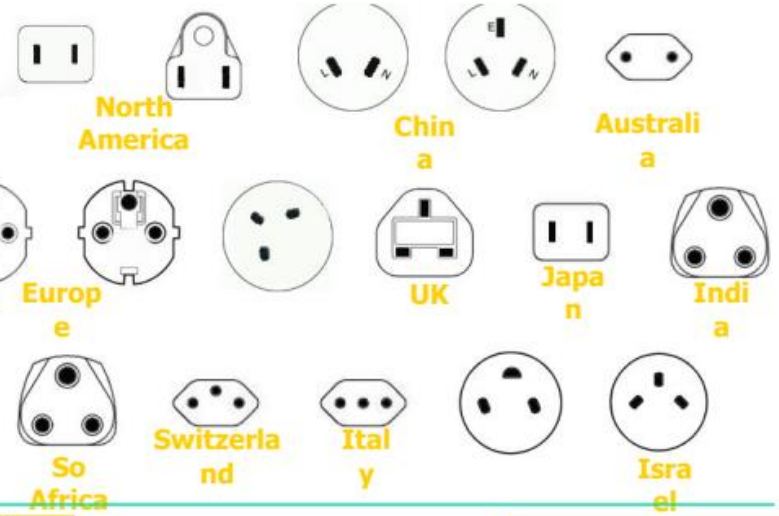
New, script based input data mapping for every study

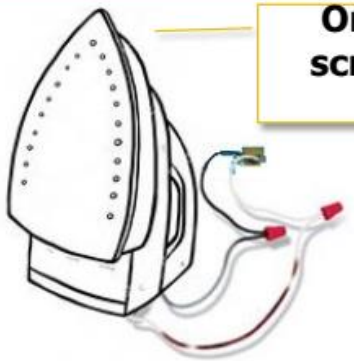
"What's the adherence to my drug in the data assets I can analyze?"



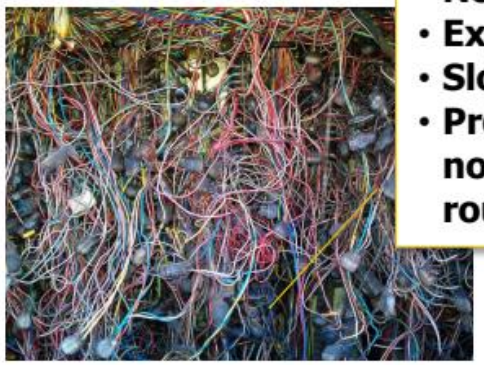
Analytical method: Adherence to Drug

Application to data



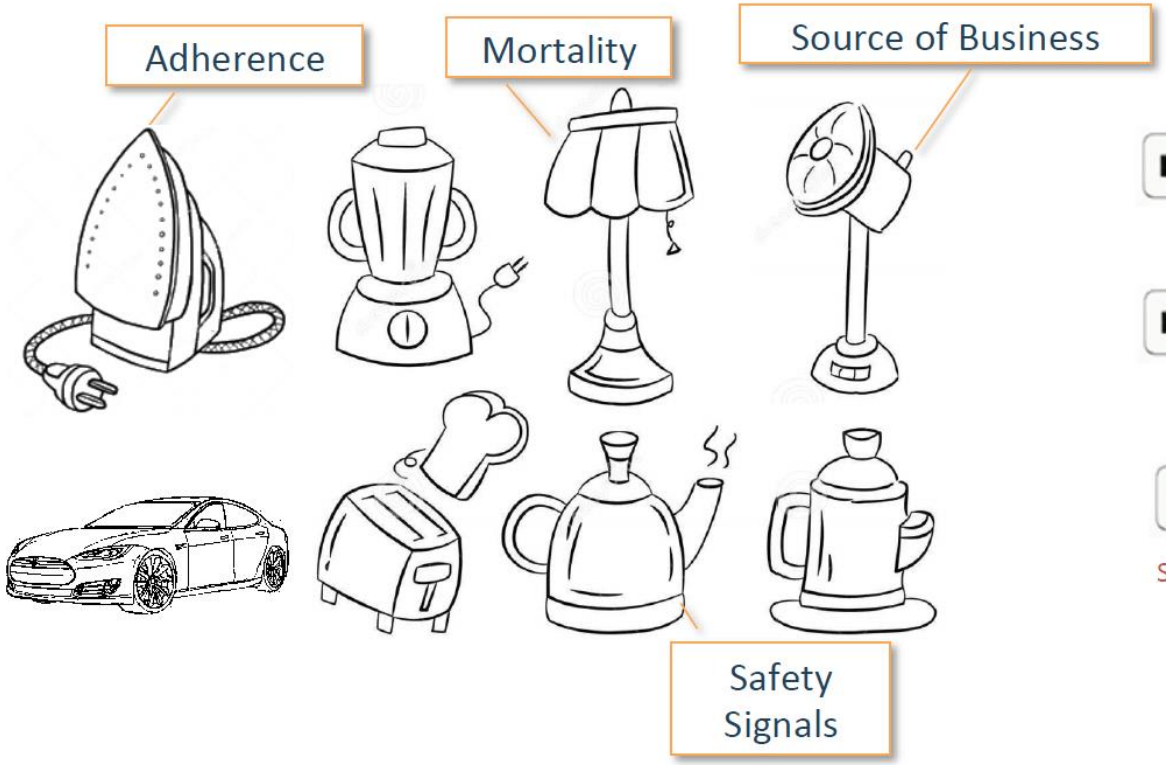


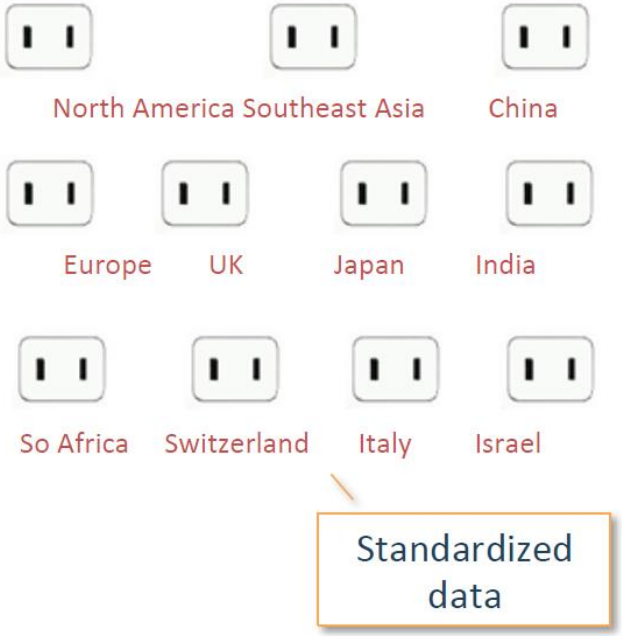
One SAS or R script for each study



- **Not scalable**
- **Expensive**
- **Slow**
- **Prohibitive to non-expert routine use**

Data standardization enables systematic research: Scalability across the World





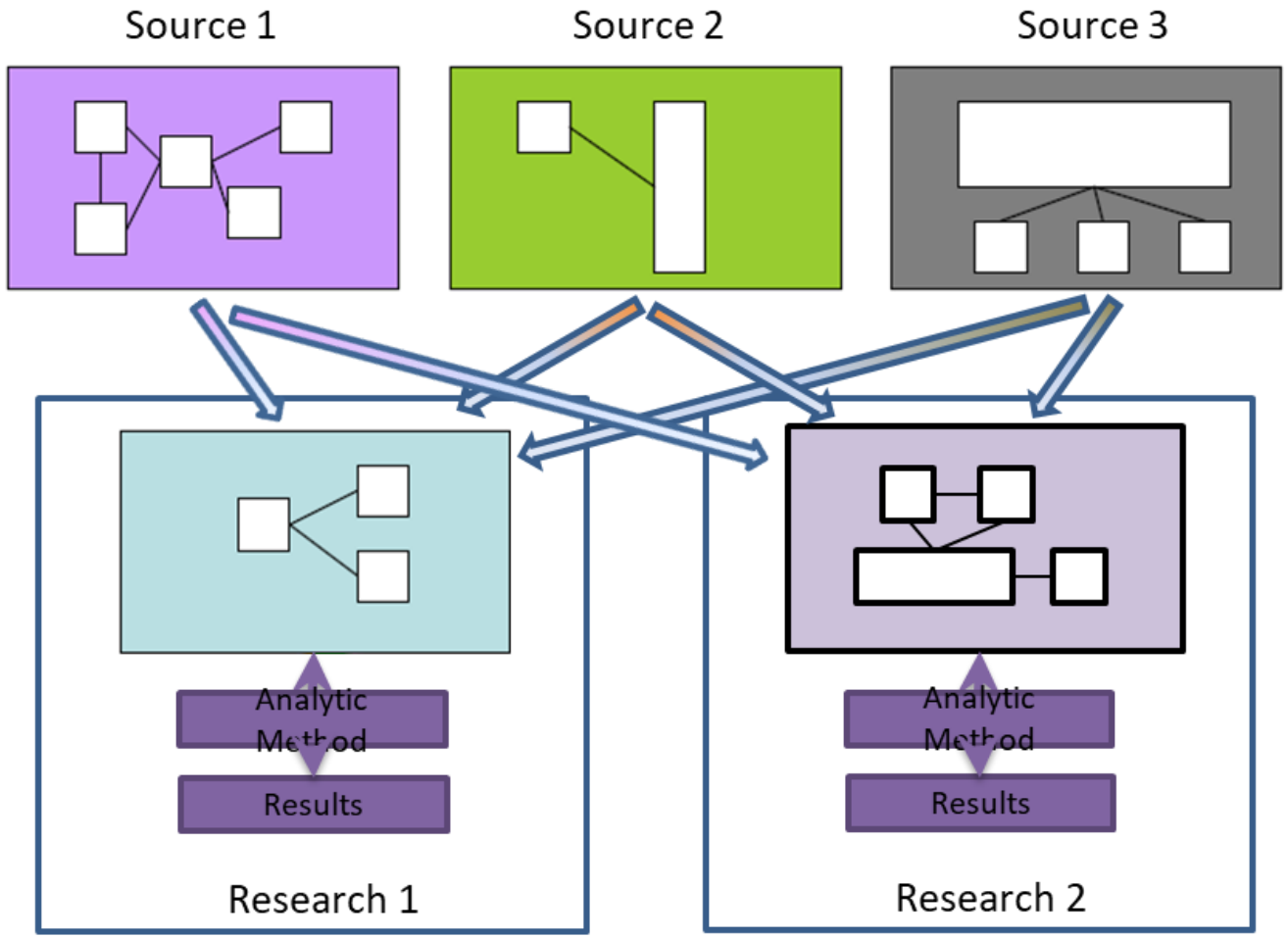
OHDSI Tools

OMOP CDM

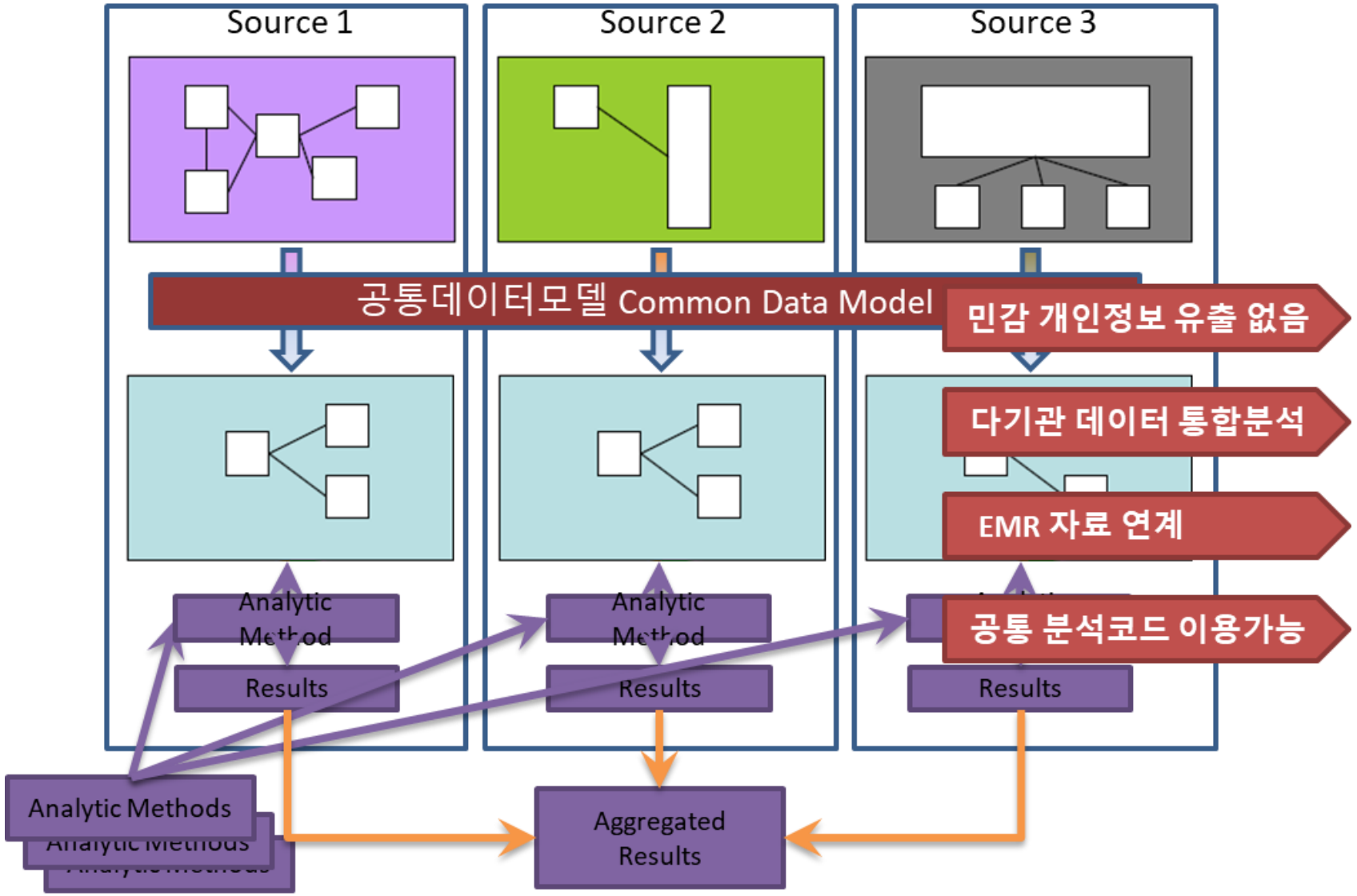
Why Common Data Model (CDM)?

기존의 다기관 연구방법

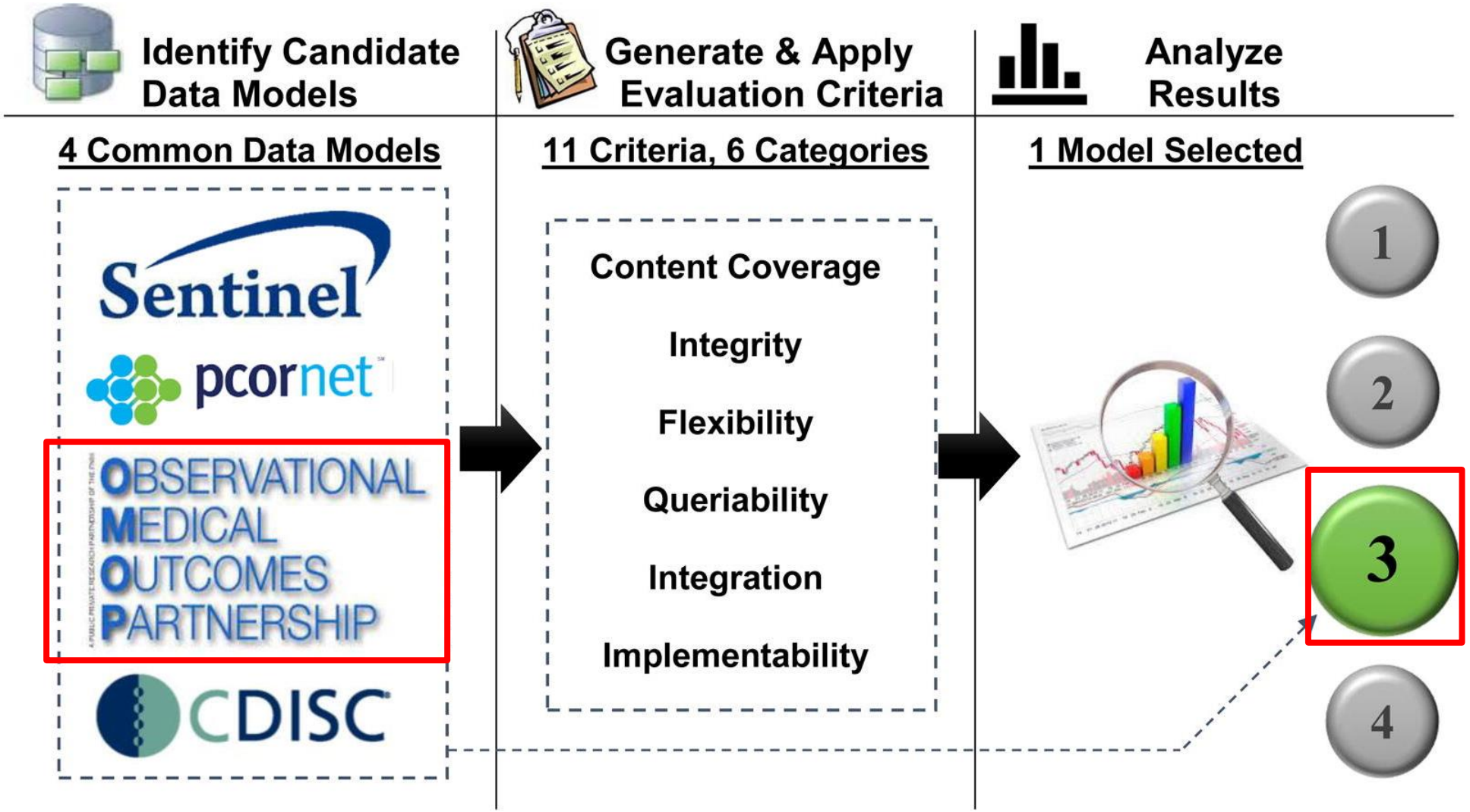
연구 수행 때마다 데이터 모델을 맞추는 변환 작업을 수행해야 함



CDM in Distributed Research Network



Various Common Data Models



- The **OMOP CDM** accommodated the highest percentage of our data elements (76%), fared well on other requirements, and had broader terminology coverage than the other models

Garza, et al., JBI 2016

OHDSI (Observational Health Data Sciences and Informatics)

- International collaborative consortium applying open-source data analytic solutions based on **OMOP-Common Data Model** (CDM) to a large network of health databases across the world

OHDSI Collaborators:

- >100 researchers in academia, industry and government
- >10 countries

OHDSI Data Network:

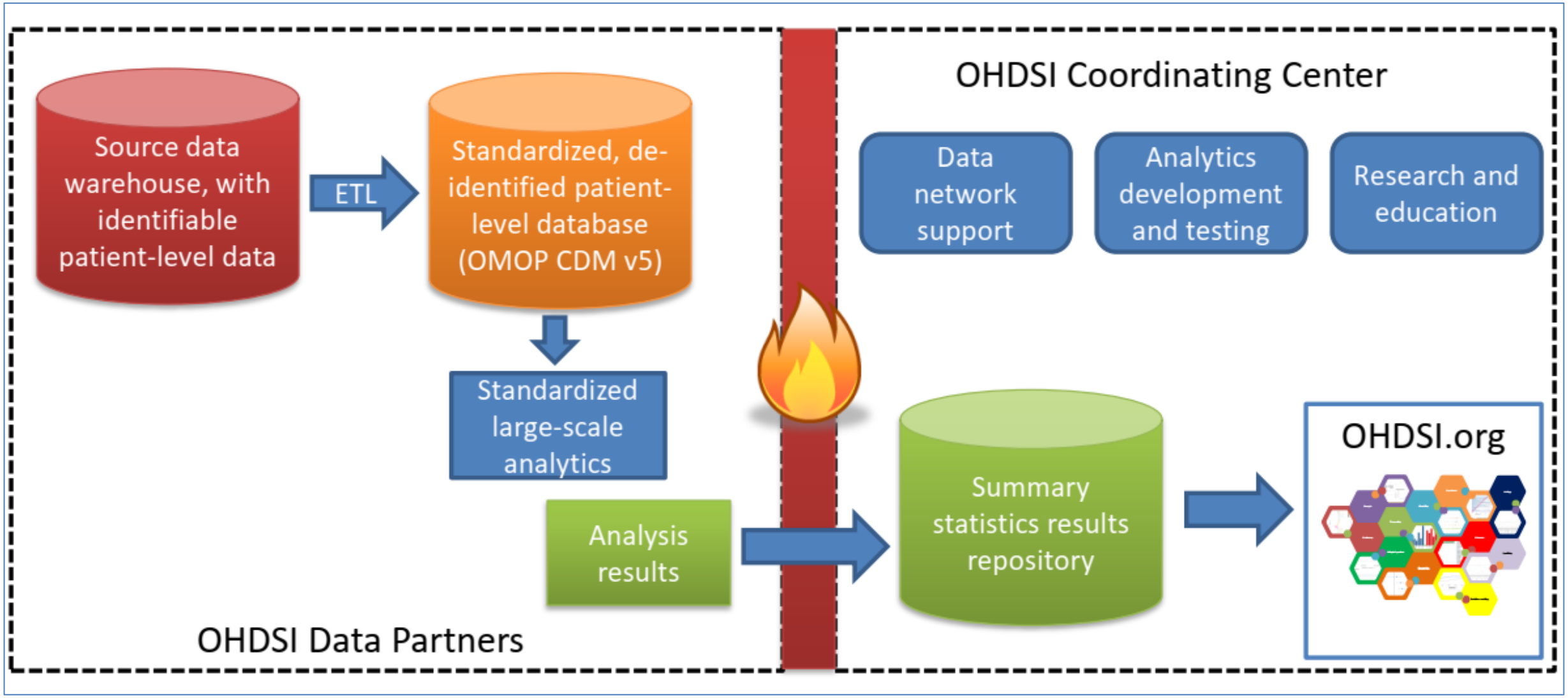
- >40 databases standardized to OMOP common data model
- >500 million patients

https://www.ohdsi.org/

Objectives of OHDSI

- **Innovation 혁신성**: Observational research is a field which will benefit greatly from disruptive thinking. We actively seek and encourage fresh methodological approaches in our work.
- **Reproducibility 재현성**: Accurate, reproducible, and well-calibrated evidence is necessary for health improvement.
- **Openness 개방성**: We strive to make all our community's proceeds open and publicly accessible, including the methods, tools and the evidence that we generate.
- **Community 공동체 정신**: Everyone is welcome to actively participate in OHDSI, whether you are a patient, a health professional, a researcher, or someone who simply believes in our cause.
- **Collaboration 협력 정신**: We work collectively to prioritize and address the real world needs of our community's participants.
- **Beneficence 선행의 정신**: We seek to protect the rights of individuals and organizations within our community at all times.

How OHDSI works



국내 CDM 데이터망 구축 현황

CDM 변환 병원 목록

No.	병원 명	병원 구분	변환 환자 수
1	가톨릭대학교 성모병원	3차	3,223,259
2	강동경희대학교병원	2차	822,183
3	강동성심병원	2차	1,662,083
4	강원대학교병원	2차	510,000
5	경북대학교병원	3차	1,002,381
6	경희의료원	3차	2,101,456
7	고려대학교 안암병원	3차	1,856,484
8	고려대학교 안산병원	3차	1,465,833
9	고려대학교 구로병원	3차	2,077,344
10	국민건강보험공단 일산병원	2차	1,358,280
11	대구가톨릭대학교병원	3차	1,688,980
12	동국대학교 일산병원	2차	779,474
13	메디플렉스 세종인천병원	2차	946,000
14	부산대학교병원	3차	1,753,002

No.	병원 명	병원 구분	변환 환자 수
15	분당서울대학교병원	3차	1,734,565
16	분당차병원	2차	2,363,386
17	서울대학교병원	3차	3,068,874
18	세종부천병원	2차	946,000
19	순천향부천병원	3차	-
20	순천향천안병원	3차	-
21	순천향구미병원	2차	-
22	순천향서울병원	2차	-
23	아주대학교병원	3차	2,400,000
24	연세원주세브란스병원	2차	-
25	원광대학교병원	3차	1,001,797
26	이화여자대학교 목동병원	2차	1,745,549
27	인하대학교병원	3차	1,977,256
28	전남대학교병원	3차	2,168,701
29	전북대학교병원	3차	1,433,023
30	칠곡경북대학교병원	3차	1,002,381
31	한양대학교병원	3차	1,783,111
32	화순전남대학교병원	3차	1,725,462

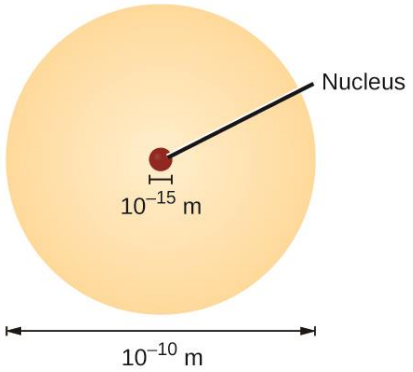
현 누적 변환 기관 수 : 32병원 (3차: 20개 / 2차: 12개)

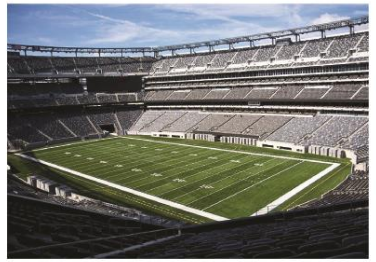
현 누적 변환 환자 수 : 44,596,864명

Phenotyping

The likelihood of transforming matter into energy
is something akin to
shooting birds in the dark in a country
where there are only a few birds.

Einstein, 1935





Phenotype

- Definition
  - A phenotype is a specification of an observable, potentially changing state of an organism, as distinguished from the genotype, which is derived from an organism's genetic makeup (*Hripcsak and Albers, JAMIA, 2018*)

What Makes a Phenotype a "Gold Standard"?

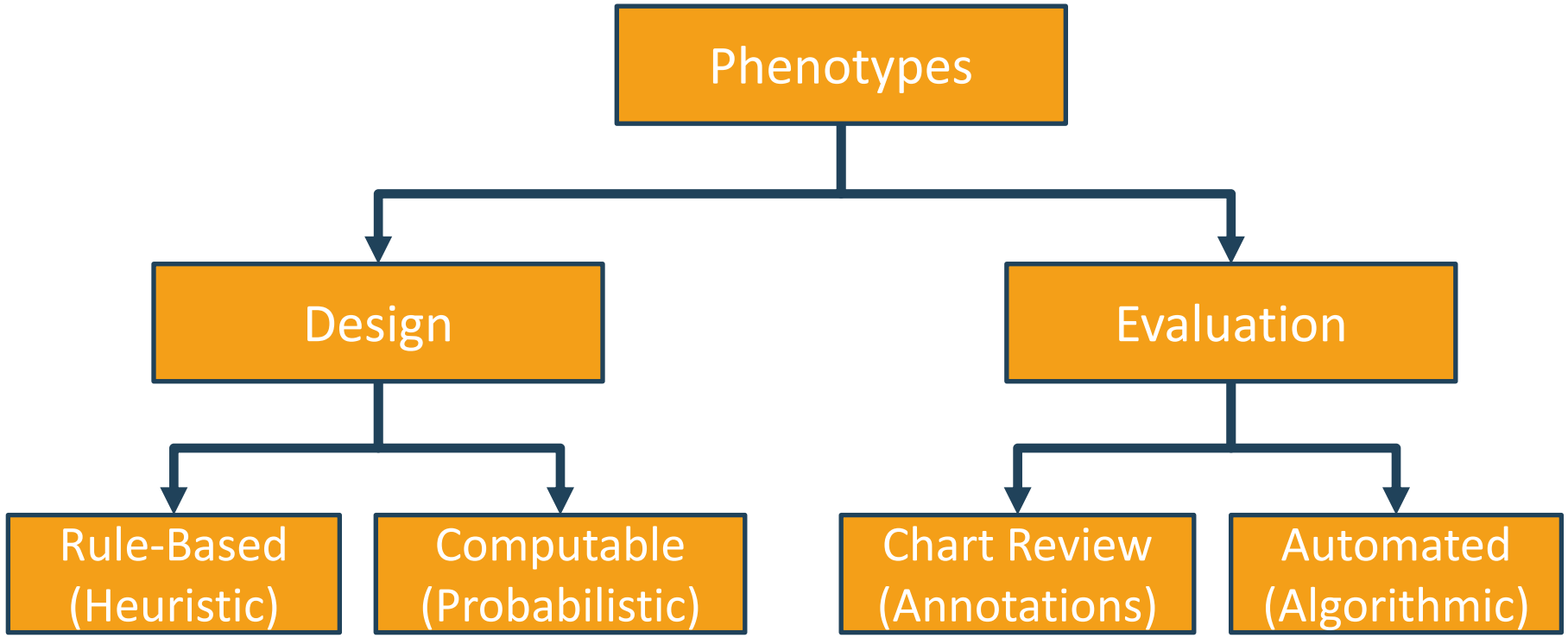
- A phenotype that has been:
  - **Designed** with best practices
  - **Evaluated** with best practices
  - **Documented** with best practices

How to define cohorts?

- Rule-based cohort definitions
  - Rule-based cohort definitions use explicit rules to describe when a patient is in the cohort.
    Defining these rules typically relies heavily on the domain expertise of the individual designing the cohort to use their knowledge of the therapeutic area of interest to build rules for cohort inclusion criteria

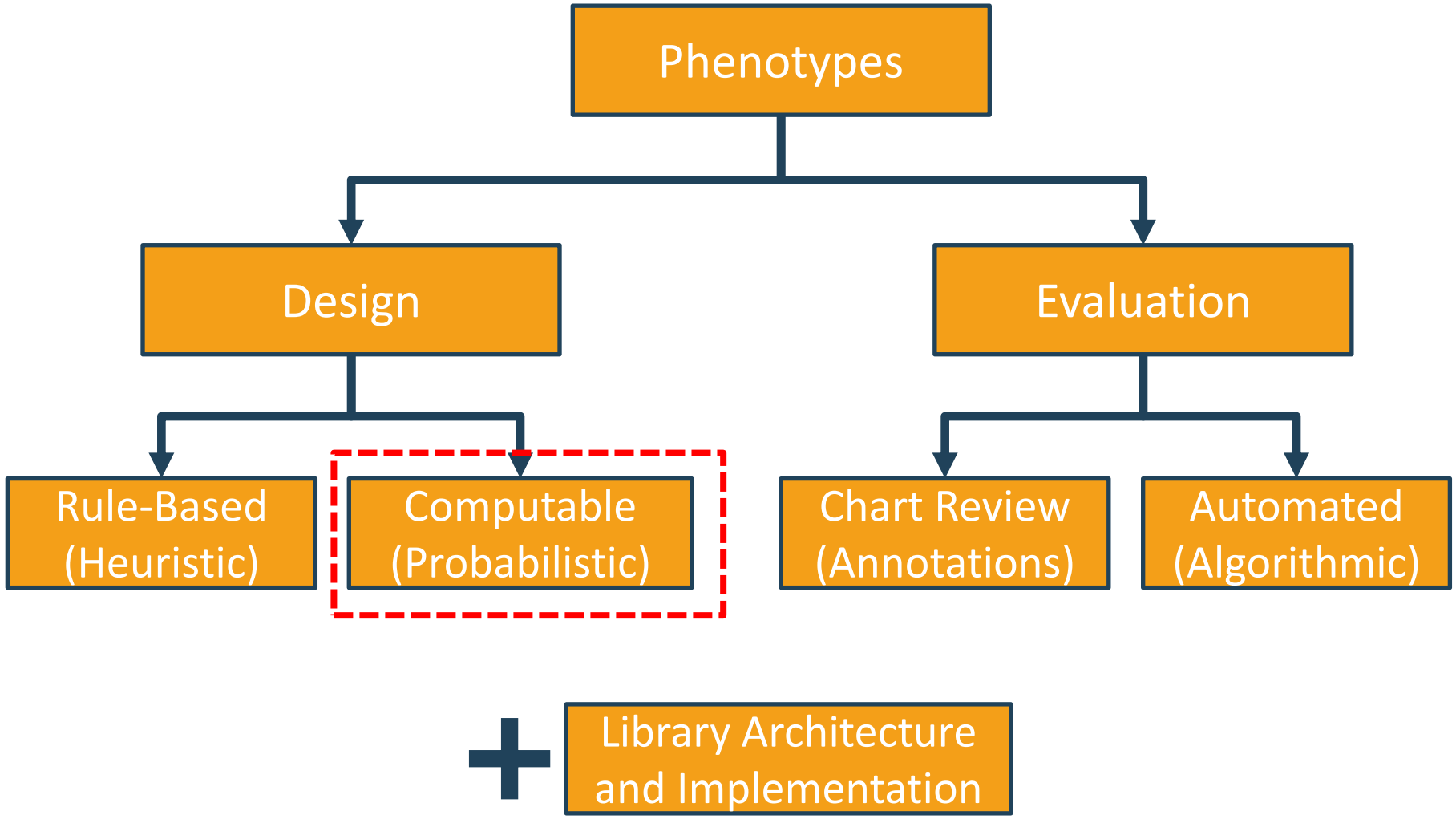
- Probabilistic cohort definitions
  - Probabilistic cohort definitions use a probabilistic model to compute a probability between 0 and 100% of the patient being in the cohort.
    This probability can be turned into a yes-no classification using some threshold, or in some study designs can be used as is. The probabilistic model is typically trained using machine learning (e.g. logistic regression) on some example data to automatically identify the relevant patient characteristics that are predictive.

Heuristic and Computational Phenotyping



+ Library Architecture and Implementation

Heuristic and Computational Phenotyping



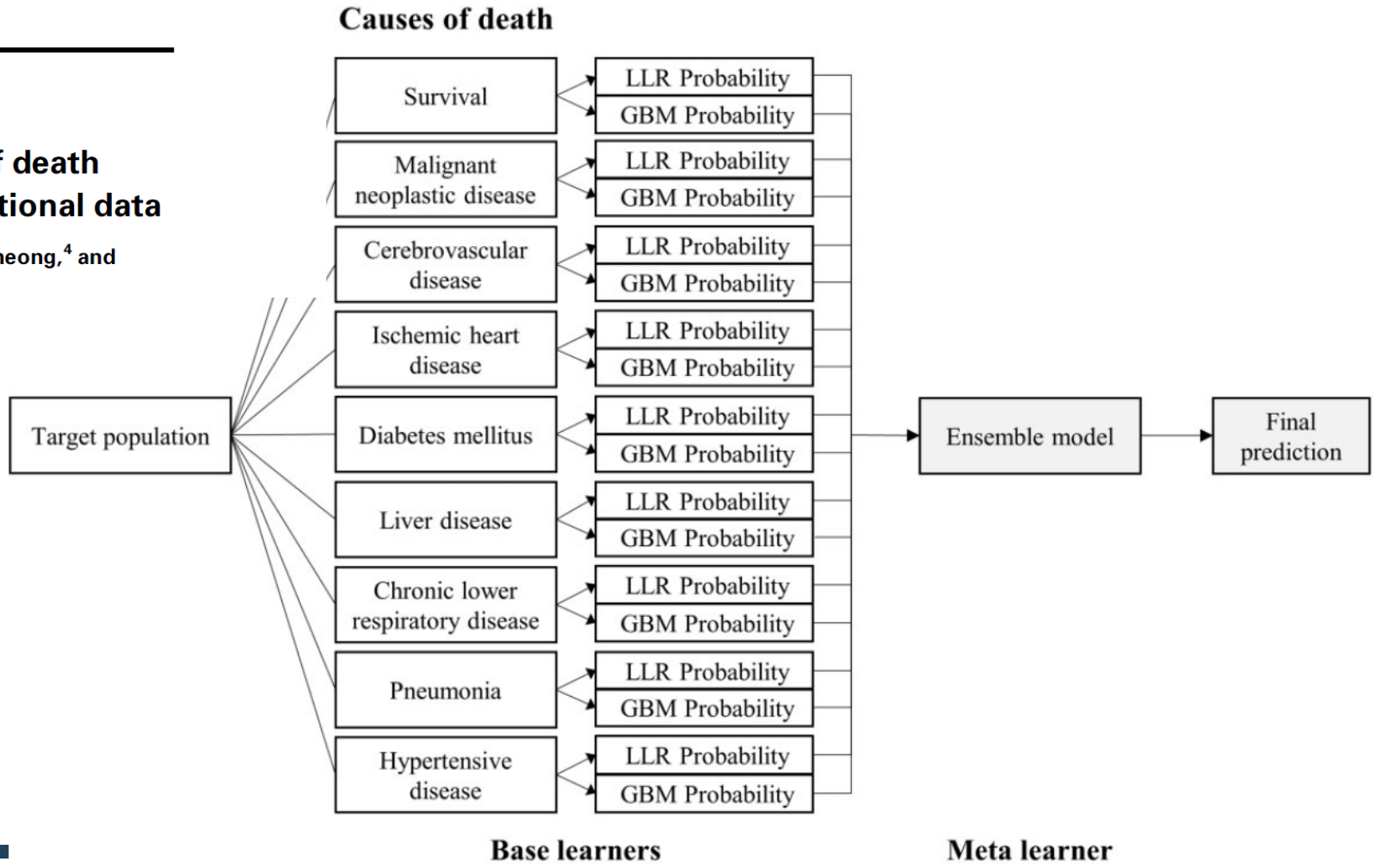
Computable phenotyping: Cause of death

Journal of the American Medical Informatics Association, 00(0), 2020, 1–11
doi: 10.1093/jamia/ocaa277
Research and Applications

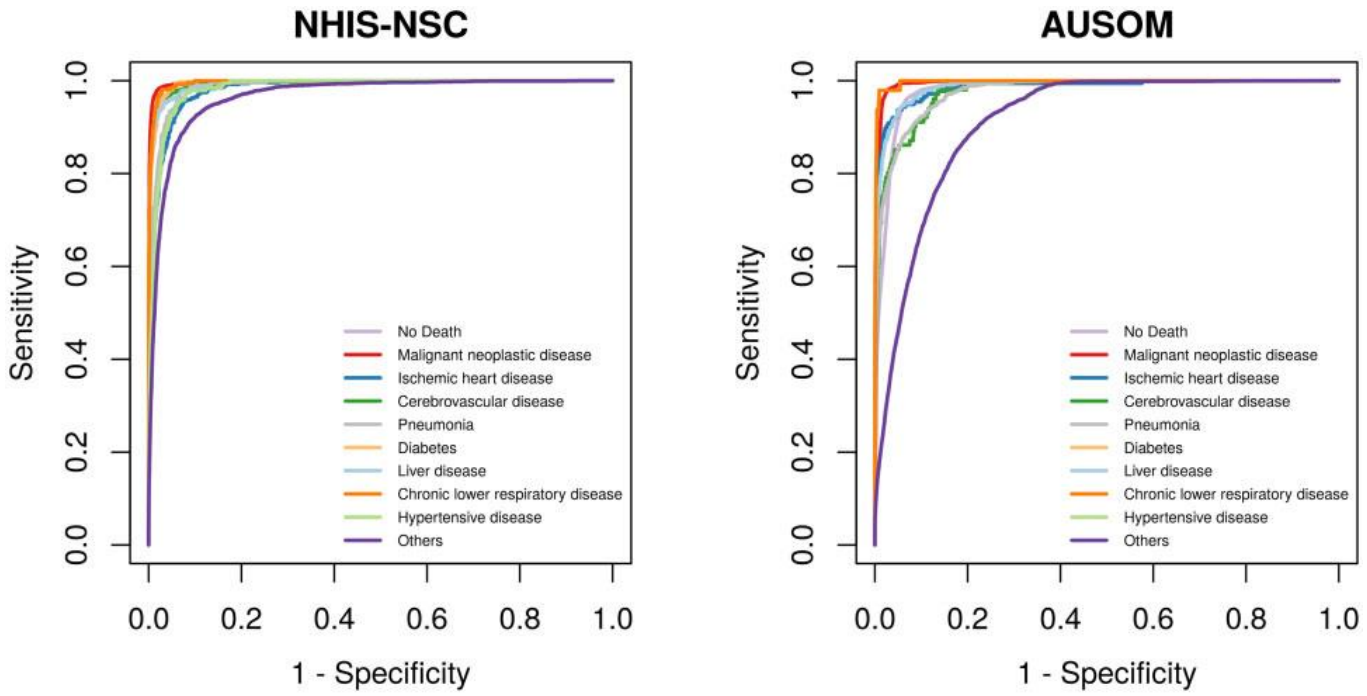
Research and Applications

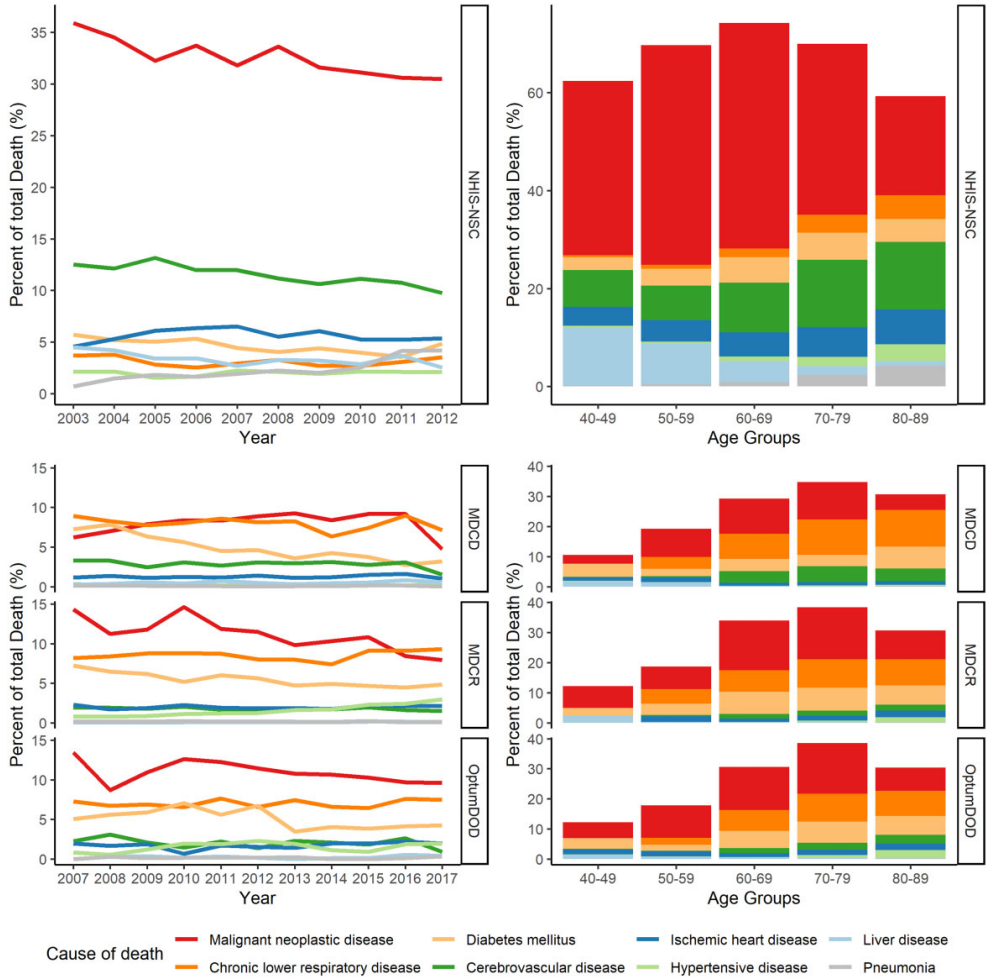
Machine-learning model to predict the cause of death using a stacking ensemble method for observational data

Chungsoo Kim [1,†] Seng Chan You,[2,†] Jenna M. Reps [3] Jae Youn Cheong,[4] and Rae Woong Park[1,2]



Computable phenotyping: Cause of death





Computable phenotyping: APHRODITE

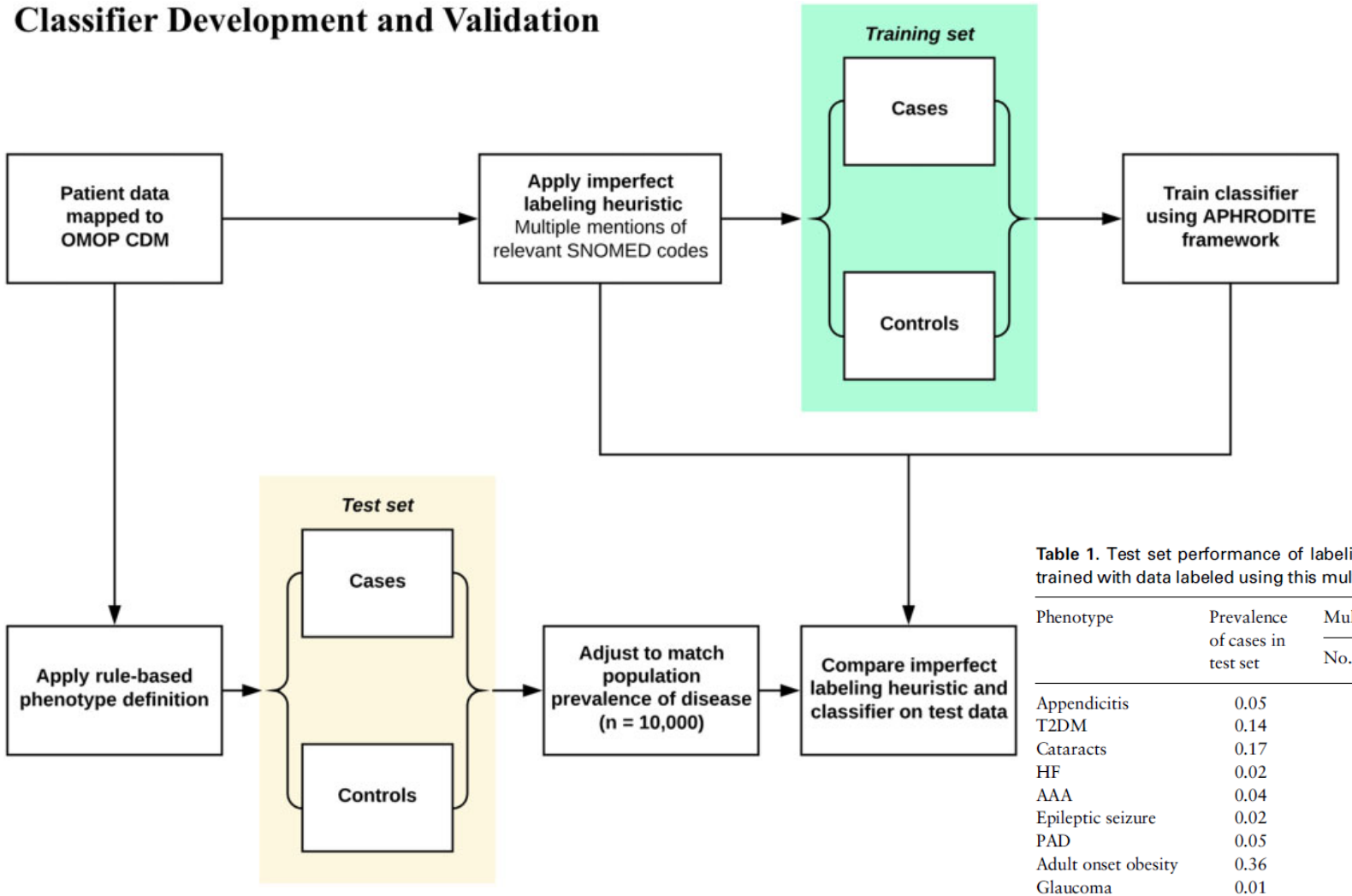
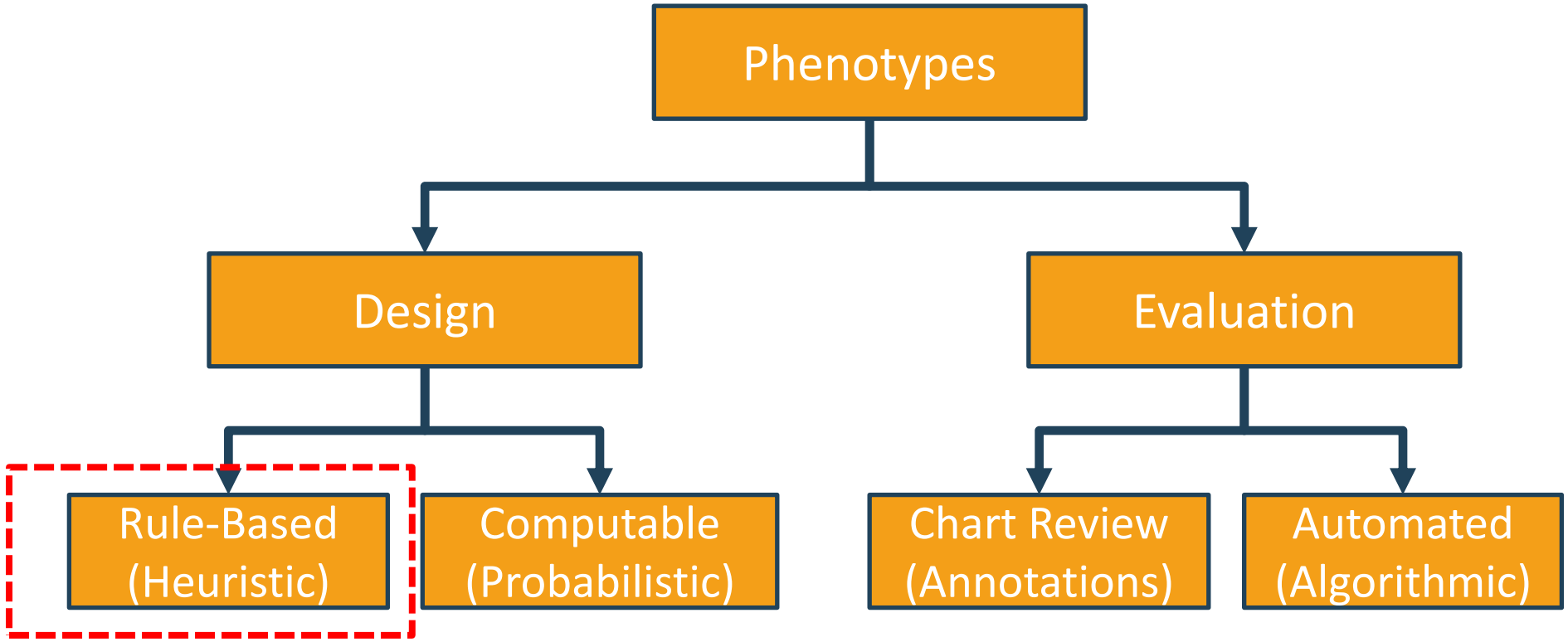


Table 1. Test set performance of labeling heuristic requiring multiple disease-specific code mentions compared to phenotype classifiers trained with data labeled using this multiple mentions approach

Phenotype	Prevalence of cases in test set	Multiple mentions of SNOMED code			APHRODITE classifier		Recall boost using classifier	Precision loss using classifier
		No. of mentions	Recall	Precision	Recall	Precision		
Appendicitis	0.05	2	0.31	1.00	0.97	0.99	0.66	0.01
T2DM	0.14	4	0.24	0.99	0.60	0.91	0.36	0.08
Cataracts	0.17	4	0.07	0.97	0.63	0.93	0.56	0.04
HF	0.02	4	0.33	0.94	0.99	0.56	0.66	0.38
AAA	0.04	4	0.22	0.99	0.53	0.97	0.31	0.02
Epileptic seizure	0.02	4	0.06	1.00	0.22	0.94	0.17	0.06
PAD	0.05	4	0.18	0.98	0.91	0.91	0.72	0.07
Adult onset obesity	0.36	4	0.20	1.00	0.29	0.91	0.09	0.09
Glaucoma	0.01	4	0.08	1.00	0.22	0.88	0.14	0.12
VTE	0.01	4	0.03	1.00	0.69	0.22	0.66	0.78

https://github.com/OHDSI/Aphrodite

Evaluation of Phenotypes



\+ Library Architecture and Implementation

EHR-derived phenotypes to claims data

Journal of Biomedical Informatics 102 (2020) 103363

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Adapting electronic health records-derived phenotypes to claims data: Lessons learned in using limited clinical data for phenotyping

Anna Ostropolets[a,d], Christian Reich[b,d], Patrick Ryan[c,d], Ning Shang[a,d], George Hripcsak[a,d,1,*], Chunhua Weng[a,d,1,*]

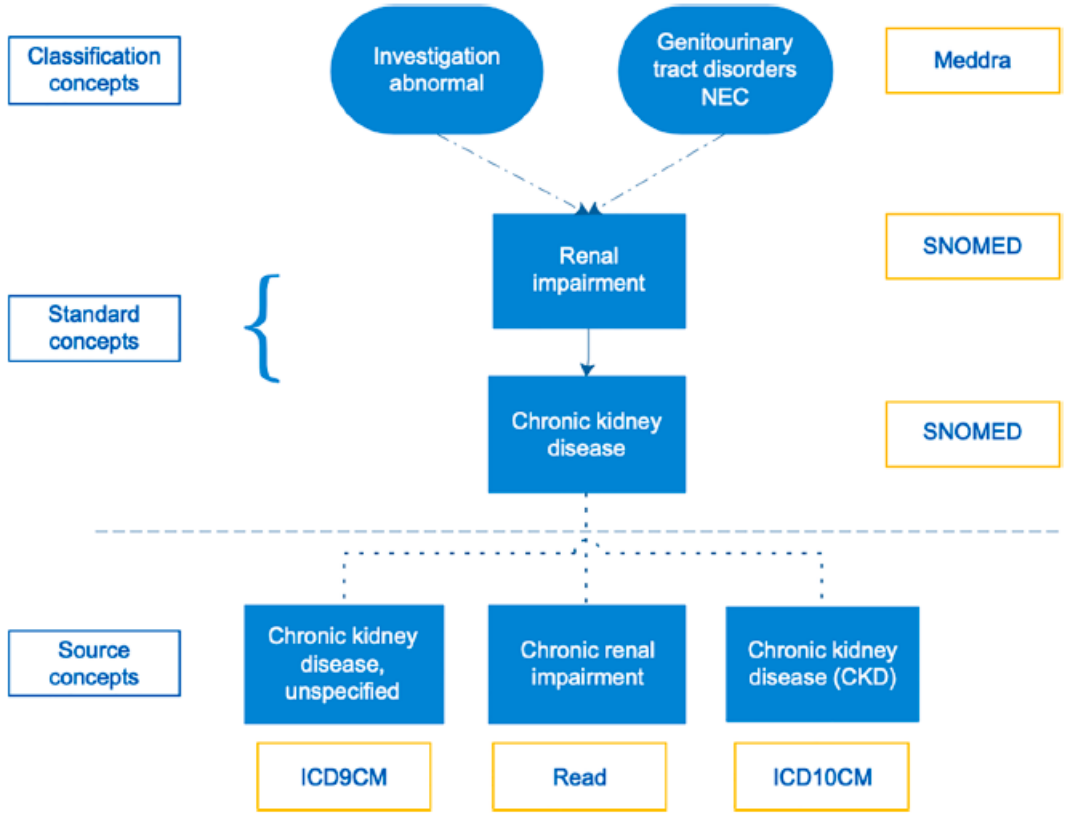


Fig. 1. General structure of OMOP Standardized Vocabularies.

Phenotyping CKD and ESRD

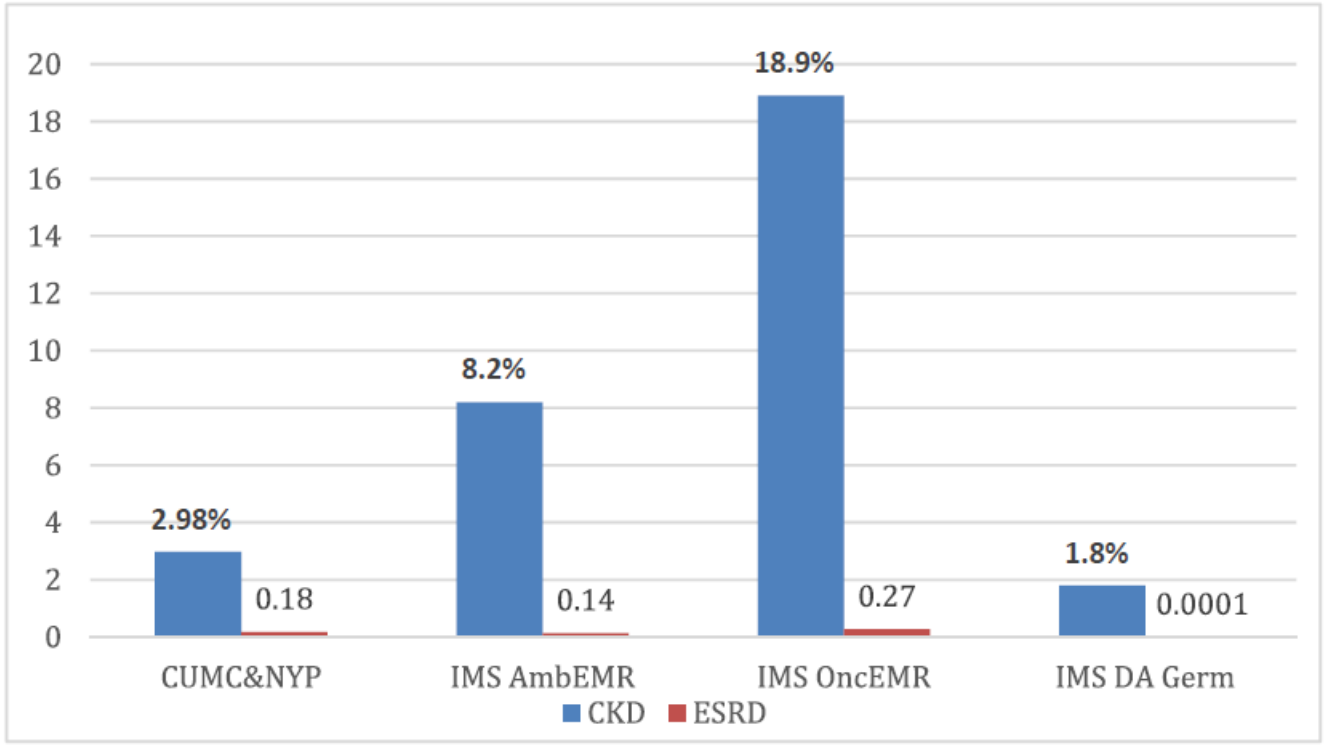


Fig. 2. Prevalence of chronic kidney disorder and end-stage renal disorder, % in four datasets.

Table 2
Positive predictive value (PPV) of the algorithms for CKD compared to the Gold Standard, %. *T – kidney transplant, D – dialysis, SKD – suspicious kidney disorders.

Dataset	Two CKD diagnosis codes	CKD or T + CKD or D + CKD	CKD or T + CKD or D + CKD or SKD + CKD
CUMC&NYP	45.1%	43.7%	41.7%
IQVIA AmbEMR	61.3%	61.6%	61.0%
IQVIA OncEMR	52.9%	62.8%	62.8%
IQVIA DA Germ	30.3%	26.0%	26.0%
Average	47.4%	48.5%	47.9%

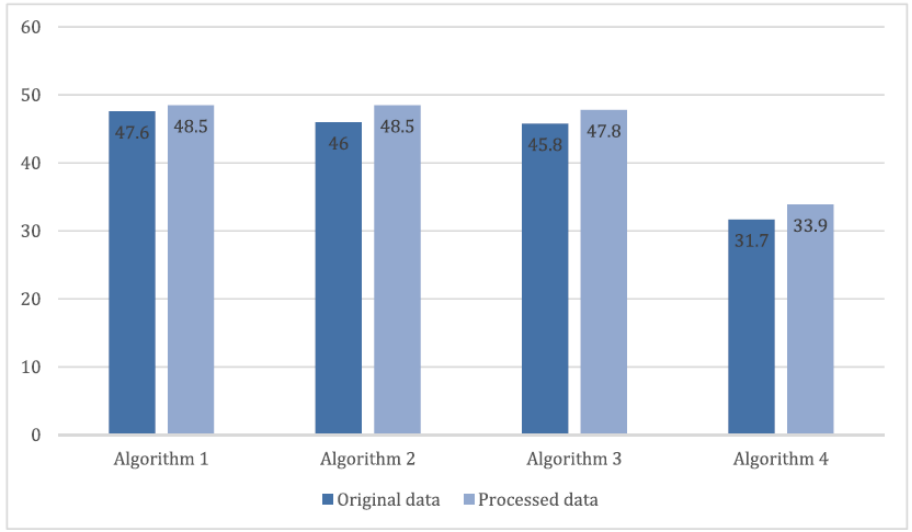


Fig. 3. Overall positive predictive value of the algorithms among four databases, %.

Heuristic and Computational Phenotyping

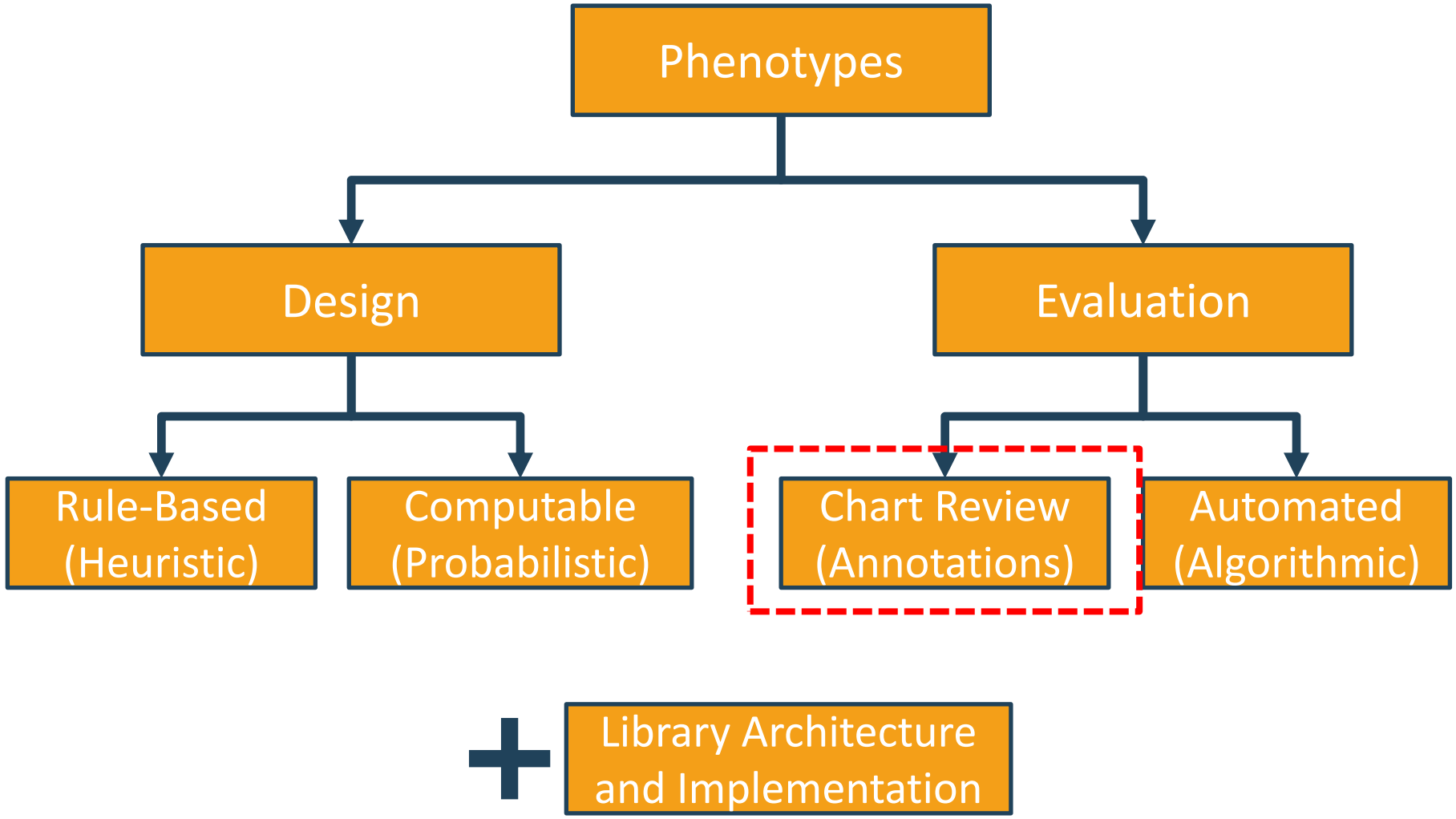
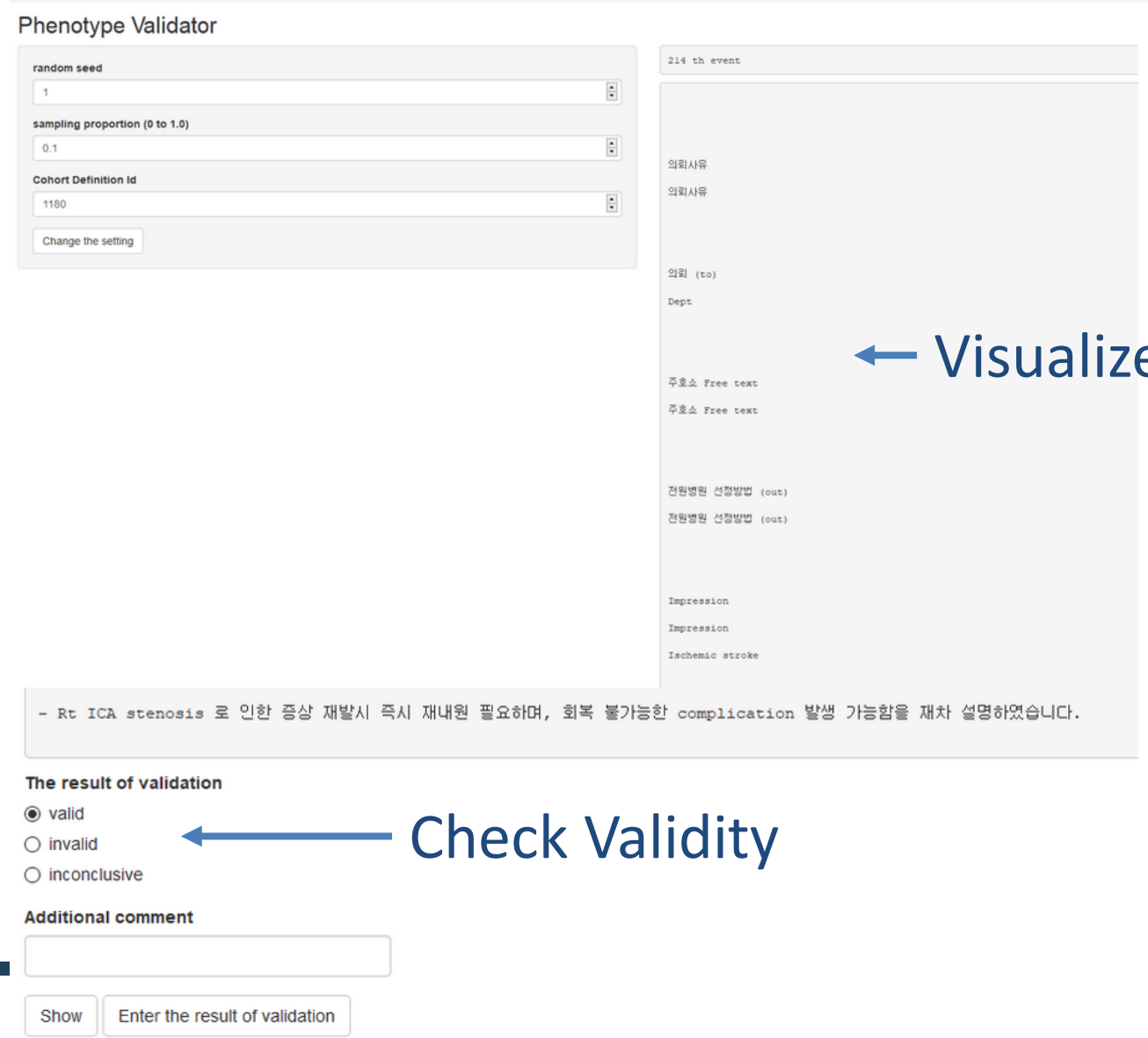


Chart Review / Evaluation

Shiny App – Seng Chan You



← Visualize Discharge Notes

← Check Validity

Evaluation of phenotypes through manual chart review

eMethod 3. Individual outcome definitions

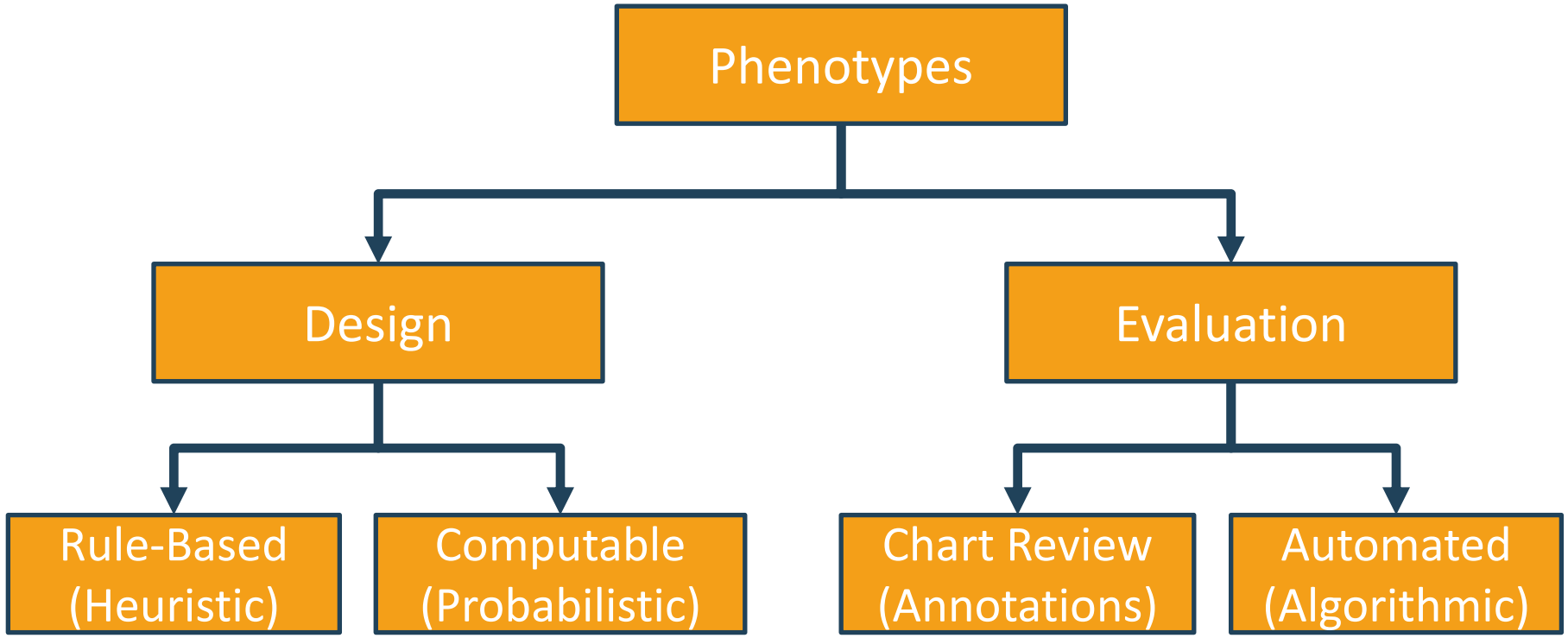
For each outcome, we developed an operational phenotype definition to determine if observational data could in fact support evaluation of the outcome. Where possible, concept sets originated with published code lists (eg ICD-9-CM and ICD-10). We developed definition of outcome cohorts and query to extract them using ATLAS, the OHDSI open-source platform (https://github.com/OHDSI/atlas). We executed these definitions on EHR data of Korean tertiary hospital to validate the definitions. Positive predictive values were estimated by a physician's manual chart review of discharge notes.

Supplementary Table. Outcome definition

Outcome	Logical description	ICD-9-CM	ICD-10	CPT4	PPV, % (n)
Acute myocardial infarction	Record of acute myocardial infarction during an inpatient or ER visit	410;410.01;410.02;410.1;410.11;410.12;410.2;410.21;410.22;410.3;410.31;410.32;410.4;410.41;410.42;410.5;410.51;410.52;410.7;410.71;410.72;410.8;410.81;410.82;410.9;410.91;410.92	I21.0;I21.1;I21.2;I21.3;I21.4;I21.9		83.8 (83/99)
Revascularization	Record of PCI or CABG during an inpatient or ER visit			566;567;33510;33511;33512;33513;33514;33516;33517;33518;33519;33521;33522;33523;33533;33534;33535;33536;33542;33545;33548;33572;33621;35506;35694;92920;92921;92924;92925;92928;92929;92933;92934;92937;92938;92941;92943;92944;1006199;1006200;1006208;1006216;1006217	100.0 (30/30)
Ischemic stroke	Earliest record of ischemic stroke during an inpatient or ER visit	346.6;346.6;346.61;346.62;346.63;433.01;433.11;433.21;433.31;433.81;433.91;434.01;434.11;434.91;997.02	I63.9;I63.8;I63.6;I63.5;I63.4;I63.3;I63.2;I63.1;I63.0;I63;G46.7;G46.6;G46.5;F01.3;F01.1;F01.0		72.9 (70/96)

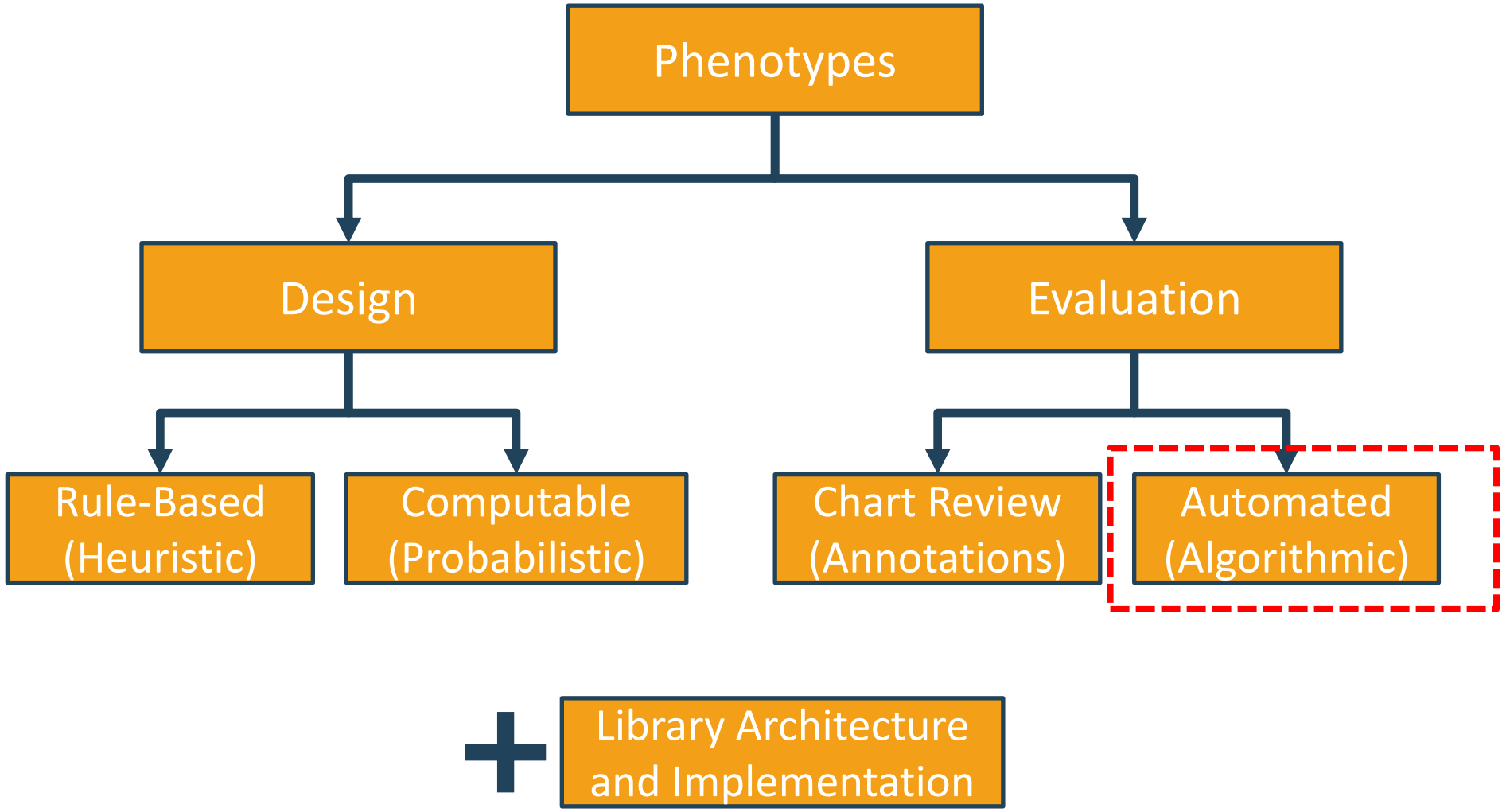
SCYou et al., JAMA, 2020

Evaluation of Phenotypes



+ Library Architecture and Implementation

Evaluation of Phenotypes



PheValuator: Phenotype algorithm evaluator

PheValuator: Development and Evaluation of a Phenotype Algorithm Evaluator.

Joel N. Swerdel, MS MPH[1,3]; Patrick B. Ryan, PhD[1,2,3]

[1] Janssen Research and Development, Titusville, NJ, USA

[2] Columbia University, New York, NY, USA

[3] Observational Health Data Sciences and Informatics (OHDSI), New York, NY

Abstract ***Background:*** *The primary approach for defining disease in observational healthcare databases is to construct phenotype algorithms (PAs), rule-based heuristics predicated on the presence, absence, and temporal logic of clinical observations. However, a complete evaluation of PAs, i.e., determining sensitivity, specificity, and positive predictive value (PPV), is rarely performed. In this study, we propose a tool (PheValuator) to perform an efficient and complete PA evaluation.* ***Method:*** *We used 4 administrative claims datasets: OptumInsight's de-identified Clinformatics™ Datamart (Eden Prairie,MN); Truven MarketScan Multi-State Medicaid); Truven MarketScan Medicare Supplemental Beneficiaries; and Truven MarketScan Commercial Claims and Encounters from 2000-2017. Using PheValuator involves 1) creating a diagnostic predictive model for the health outcome of interest (HOI), 2) applying the model to a large set of randomly selected subjects, and 3) comparing each subject's HOI predicted probability to inclusion/exclusion in PAs. We used the predicted probability as a measure to classify positive/negative cases. We examined 4 HOI's: myocardial infarction, cerebral infarction, diabetes mellitus, and atrial fibrillation. We examined several PAs for each HOI, including 1-time (1X) occurrence of the diagnosis code in the subject's record and 1-time occurrence of the diagnosis in an inpatient setting with the HOI code as the primary reason for admission (1X-IP-1stPos).* ***Results:*** *Across HOIs, the 1X PA showed the highest sensitivity/lowest PPV among all PAs. 1X-IP-1stPos yielded the highest PPV/lowest sensitivity. Specificity was very high across algorithms. We found similar results between algorithms across datasets.* ***Conclusion:*** *PheValuator appears to show promise as a tool to assess PA performance characteristics.*

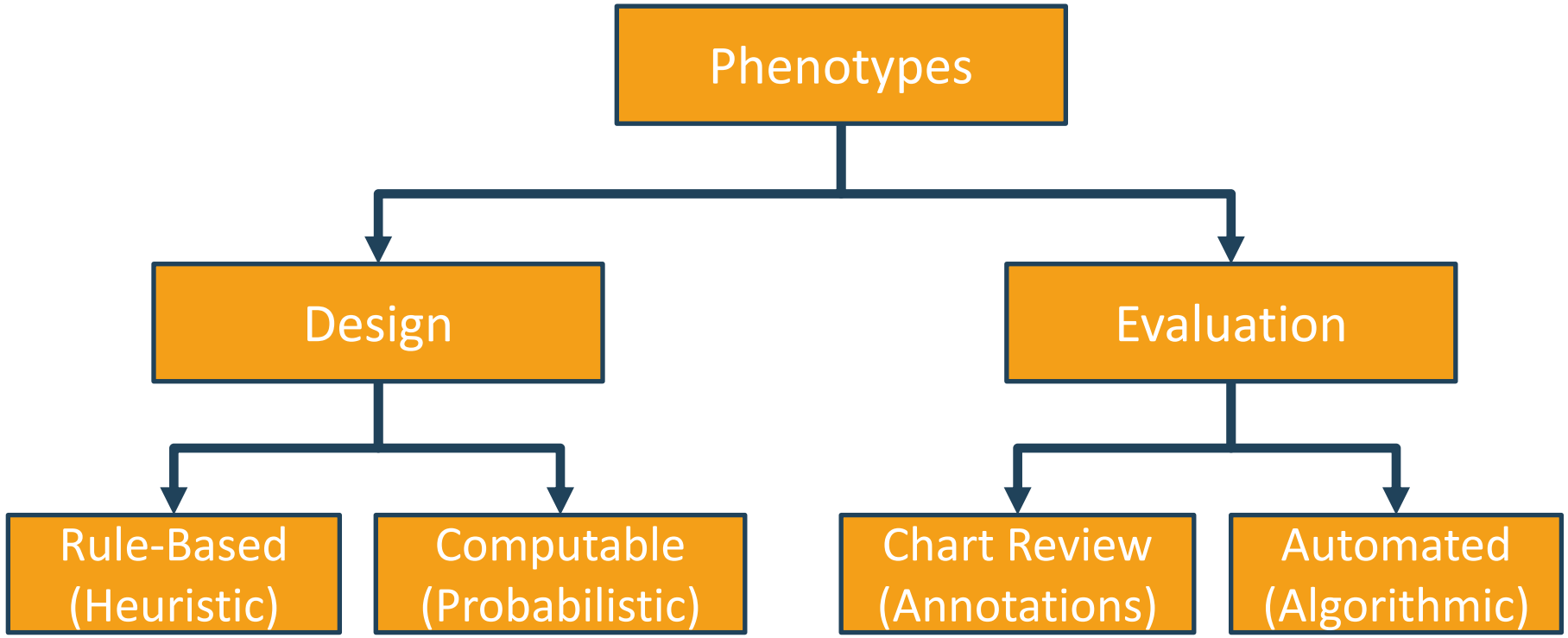
PheValuator: Phenotype algorithm evaluator

Table 1: Performance Evaluation of Phenotype Algorithms using PheValuator

Phenotype Algorithm	CDM	Myocardial Infarction			Cerebral Infarction			Diabetes Mellitus			Atrial Fibrillation		
		Sens	PPV	Spec	Sens	PPV	Spec	Sens	PPV	Spec	Sens	PPV	Spec
>=1 x HOI	CCAE	99.2%	81.9%	99.9%	99.1%	66.6%	99.8%	99.7%	73.7%	97.2%	99.9%	82.7%	99.8%
	Optum	99.9%	86.6%	99.8%	99.9%	71.0%	99.4%	99.6%	86.4%	97.9%	99.9%	87.6%	99.5%
	MDCD	99.8%	73.5%	99.3%	99.2%	66.3%	98.5%	99.7%	80.4%	95.6%	99.9%	82.8%	99.2%
	MDCR	99.7%	84.4%	99.0%	99.7%	70.0%	97.5%	99.7%	79.6%	92.6%	99.9%	85.3%	97.2%
>=2 x HOI	CCAE	61.9%	91.0%	99.9%	55.7%	80.3%	99.9%	82.9%	80.3%	98.4%	72.9%	86.6%	99.9%
	Optum	54.6%	91.9%	99.9%	56.7%	81.7%	99.8%	81.4%	89.2%	98.7%	78.3%	90.1%	99.7%
	MDCD	54.5%	81.3%	99.8%	60.7%	75.7%	99.4%	83.0%	83.7%	97.0%	71.2%	85.1%	99.5%
	MDCR	54.4%	88.5%	99.6%	58.4%	80.5%	99.2%	84.4%	82.2%	94.7%	79.6%	87.6%	98.2%
>=1 x HOI - In-Patient	CCAE	79.3%	89.1%	99.9%	59.6%	77.7%	99.9%	30.3%	84.5%	99.6%	40.0%	89.4%	99.9%
	Optum	75.3%	91.5%	99.9%	60.7%	83.6%	99.8%	39.5%	92.1%	99.6%	54.4%	91.5%	99.8%
	MDCD	73.1%	78.6%	99.6%	56.4%	75.4%	99.5%	63.7%	85.1%	98.0%	59.5%	88.3%	99.7%
	MDCR	75.6%	87.5%	99.4%	63.7%	78.6%	99.0%	41.1%	83.1%	97.6%	59.1%	87.9%	98.7%
>=1 x HOI, IP - 1st Position	CCAE	72.6%	90.3%	99.9%	55.0%	79.6%	99.9%	6.9%	88.1%	99.9%	30.3%	90.3%	99.9%
	Optum	62.5%	92.5%	99.9%	54.7%	87.4%	99.9%	7.4%	93.6%	99.9%	33.4%	93.1%	99.9%
	MDCD	57.2%	84.3%	99.8%	47.7%	78.0%	99.6%	14.6%	86.6%	99.6%	29.7%	93.4%	99.9%
	MDCR	65.2%	88.3%	99.5%	58.9%	80.0%	99.2%	13.0%	83.5%	99.3%	42.4%	88.9%	99.1%

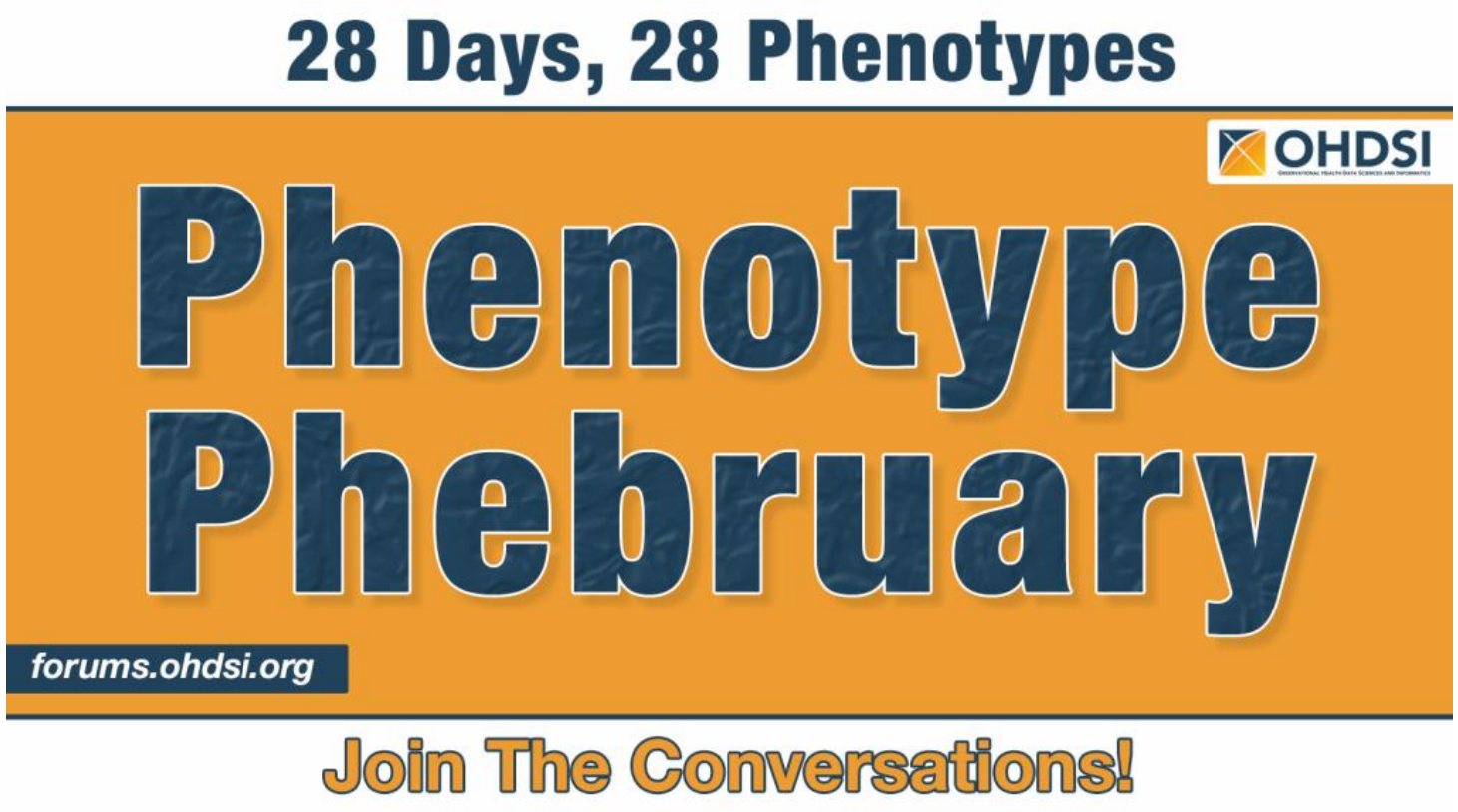
CDM - Common Data Model; Sens - Sensitivity; PPV - Positive Predictive Value; Spec - Specificity; CCAE - Truven MarketScan Commercial Claims and Encounters; MDCD - Truven Medicaid; MDCR - Truven Medicare; Optum -

Evaluation of Phenotypes



+ Library Architecture and Implementation

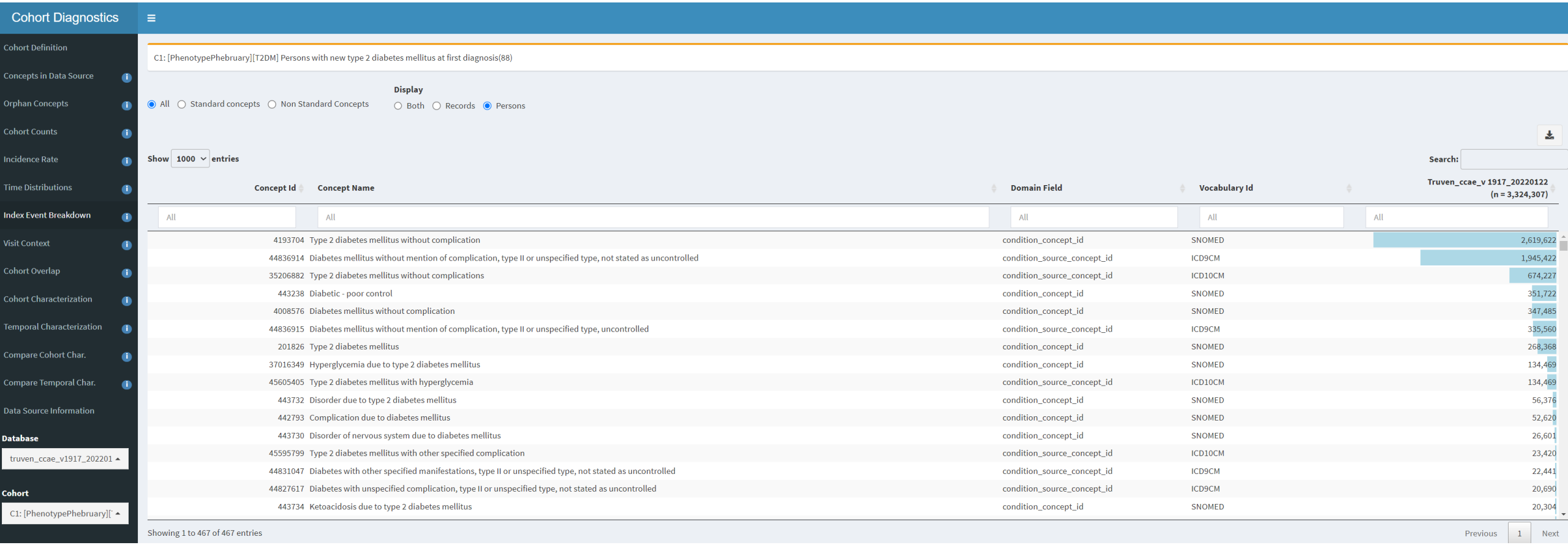
Heritage of Phenotype Phebruary



Feb. 1 • Type 2 Diabetes Mellitus
Feb. 2 • Type 1 Diabetes Mellitus
Feb. 3 • Atrial Fibrillation
Feb. 4 • Multiple Myeloma
Feb. 5 • Alzheimer's Disease
Feb. 6 • Hemorrhagic Events
Feb. 7 • Neutropenia
Feb. 8 • Kidney Stones
Feb. 9 • Delirium
Feb. 10 • Systemic Lupus Erythematosus
Feb. 11 • Suicide Attempts
Feb. 12 • Parkinson's Disease and Parkinsonism
Feb. 13 • Attention Deficit Hyperactivity Disorder
Feb. 14 • Hypertension *(Video Description)*
Feb. 15 • Acute Myocardial Infarction
Feb. 16 • Heart Failure
Feb. 17 • Cardiomyopathy
Feb. 18 • Multiple Sclerosis
Feb. 19 • Triple Negative Breast Cancer
Feb. 20 • Pulmonary Hypertension
Feb. 21 • Prostate Cancer
Feb. 22 • HIV
Feb. 23 • Hidradenitis Suppurativa
Feb. 24 • Anaphylaxis
Feb. 25 • Depression
Feb. 26 • Non-Small-Cell Lung Cancer
Feb. 27 • Drug-Induced Liver Injury
Feb. 28 • Severe Visual Impairment And Blindness
Bonus • Acute Kidney Injury

https://www.ohdsi.org/phenotype-phebruary/

Evaluating phenotypes through cohort diagnostics



https://data.ohdsi.org/phenotypePhebruary/

Effect of vocabulary mapping for conditions on phenotype cohorts

Journal of the American Medical Informatics Association, 0(0), 2018, 1–8
doi: 10.1093/jamia/ocy124
Research and Applications

Research and Applications

Effect of vocabulary mapping for conditions on phenotype cohorts

George Hripcsak,[1,2,3] Matthew E Levine,[1,2] Ning Shang,[1,2] and Patrick B Ryan[1,2,4]

[1]Department of Biomedical Informatics, Columbia University, New York, New York, USA, [2]Observational Health Data Sciences and Informatics (OHDSI), New York, New York, USA, [3]Medical Informatics Services, NewYork-Presbyterian Hospital, New York, New York, USA, and [4]Epidemiology Analytics, Janssen Research and Development, Titusville, New Jersey, USA

Corresponding Author: George Hripcsak, MD, MS, Department of Biomedical Informatics, Columbia University Irving Medical Center, 622 W 168th St, PH20, New York, NY 10032, USA (hripcsak@columbia.edu)

Received 27 April 2018; Revised 13 August 2018; Editorial Decision 22 August 2018; Accepted 3 September 2018

ABSTRACT

Objective: To study the effect on patient cohorts of mapping condition (diagnosis) codes from source billing vocabularies to a clinical vocabulary.

Materials and Methods: Nine International Classification of Diseases, Ninth Revision, Clinical Modification (ICD9-CM) concept sets were extracted from eMERGE network phenotypes, translated to Systematized Nomenclature of Medicine - Clinical Terms concept sets, and applied to patient data that were mapped from source ICD9-CM and ICD10-CM codes to Systematized Nomenclature of Medicine - Clinical Terms codes using Observational Health Data Sciences and Informatics (OHDSI) Observational Medical Outcomes Partnership (OMOP) vocabulary mappings. The original ICD9-CM concept set and a concept set extended to ICD10-CM were used to create patient cohorts that served as gold standards.

Results: Four phenotype concept sets were able to be translated to Systematized Nomenclature of Medicine - Clinical Terms without ambiguities and were able to perform perfectly with respect to the gold standards. The other 5 lost performance when 2 or more ICD9-CM or ICD10-CM codes mapped to the same Systematized Nomenclature of Medicine - Clinical Terms code. The patient cohorts had a total error (false positive and false negative) of up to 0.15% compared to querying ICD9-CM source data and up to 0.26% compared to querying ICD9-CM and ICD10-CM data. Knowledge engineering was required to produce that performance; simple automated methods to generate concept sets had errors up to 10% (one outlier at 250%).

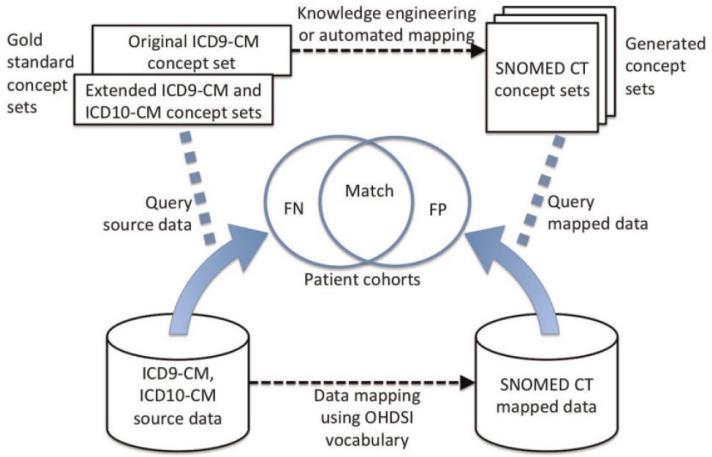
Discussion: The translation of data from source vocabularies to Systematized Nomenclature of Medicine - Clinical Terms (SNOMED CT) resulted in very small error rates that were an order of magnitude smaller than other error sources.

Conclusion: It appears possible to map diagnoses from disparate vocabularies to a single clinical vocabulary and carry out research using a single set of definitions, thus improving efficiency and transportability of research.

error sources.

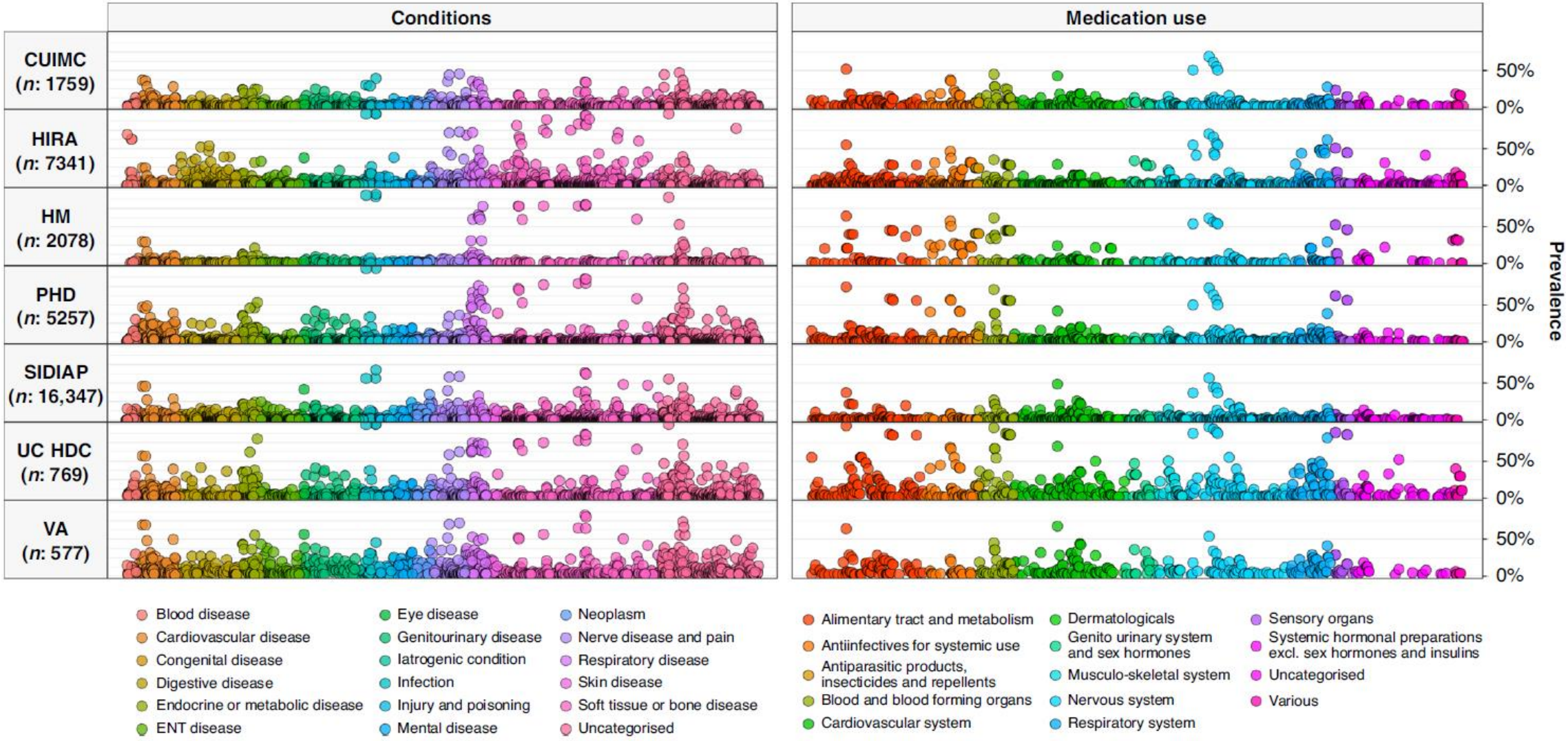
Conclusion: It appears possible to map diagnoses from disparate vocabularies to a single clinical vocabulary and carry out research using a single set of definitions, thus improving efficiency and transportability of research.

Effect of vocabulary mapping for conditions on phenotype cohorts



Pheno	#Cases	Original		Knowledge engineered				Automated concept set creation							
		ICD9 set[a]		SNOMED mimic		SNOMED optimize		SNOMED no desc		SNOMED all desc		SNOMED desc x child		SNOMED desc x desc	
		FP	FN	FP	FN	FP	FN	FP	FN	FP	FN	FP	FN	FP	FN
HF	75 312	0	0	0	0	0	0	0	0	1262	0	1054	0	1054	0
HF2	75 312	0	0	0	0	0	0	0	0	1262	0	1054	0	1054	0
T1DM	27 861	0	0	0	23	0	23	108	0	943	0	943	0	108	0
T2DM	125 342	0	0	3	30	3	30	34	0	1318	0	104	0	34	0
Appy	9887	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ADHD	14 399	0	0	0	19	0	19	1362	0	1362	0	1362	0	1362	0
Catar	50 879	0	0	50	0	74	0	50	0	2491	0	80	0	80	0
Crohn	4679	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RA	9655	0	0	0	0	0	0	0	0	25 103	0	0	0	0	0

Deep phenotyping of 34,128 patients hospitalized with COVID-19 in an international network study



EBurn, SCYou, et al., Nature Communications, 2020

Remained Variations in Phenotypes across the databases

RESEARCH: SPECIAL PAPER

FAST TRACK

Characterising the background incidence rates of adverse events of special interest for covid-19 vaccines in eight countries: multinational network cohort study

Xintong Li,[1] Anna Ostropolets,[2] Rupa Makadia,[3] Azza Shoaibi,[3] Gowtham Rao,[3] Anthony G Sena,[3,6] Eugenia Martinez-Hernandez,[4] Antonella Delmestri,[1] Katia Verhamme,[6,7] Peter R Rijnbeek,[6] Talita Duarte-Salles,[5] Marc A Suchard,[8,9] Patrick B Ryan,[2,3] George Hripcsak,[2] Daniel Prieto-Alhambra[1,6]

For numbered affiliations see end of the article.

Correspondence to:
D Prieto-Alhambra Botnar Research Centre, Oxford, UK
daniel.prietoalhambra@ndorms.ox.ac.uk
(or@prieto_alhambra on Twitter:
ORCID 0000-0002-3950-6346)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2021;373:n1435
http://dx.doi.org/10.1136/bmj.n1435

Accepted: 3 June 2021

ABSTRACT

OBJECTIVE

To quantify the background incidence rates of 15 prespecified adverse events of special interest (AESIs) associated with covid-19 vaccines.

DESIGN

Multinational network cohort study.

SETTING

Electronic health records and health claims data from eight countries: Australia, France, Germany, Japan, the Netherlands, Spain, the United Kingdom, and the United States, mapped to a common data model.

PARTICIPANTS

126 661 070 people observed for at least 365 days before 1 January 2017, 2018, or 2019 from 13 databases.

MAIN OUTCOME MEASURES

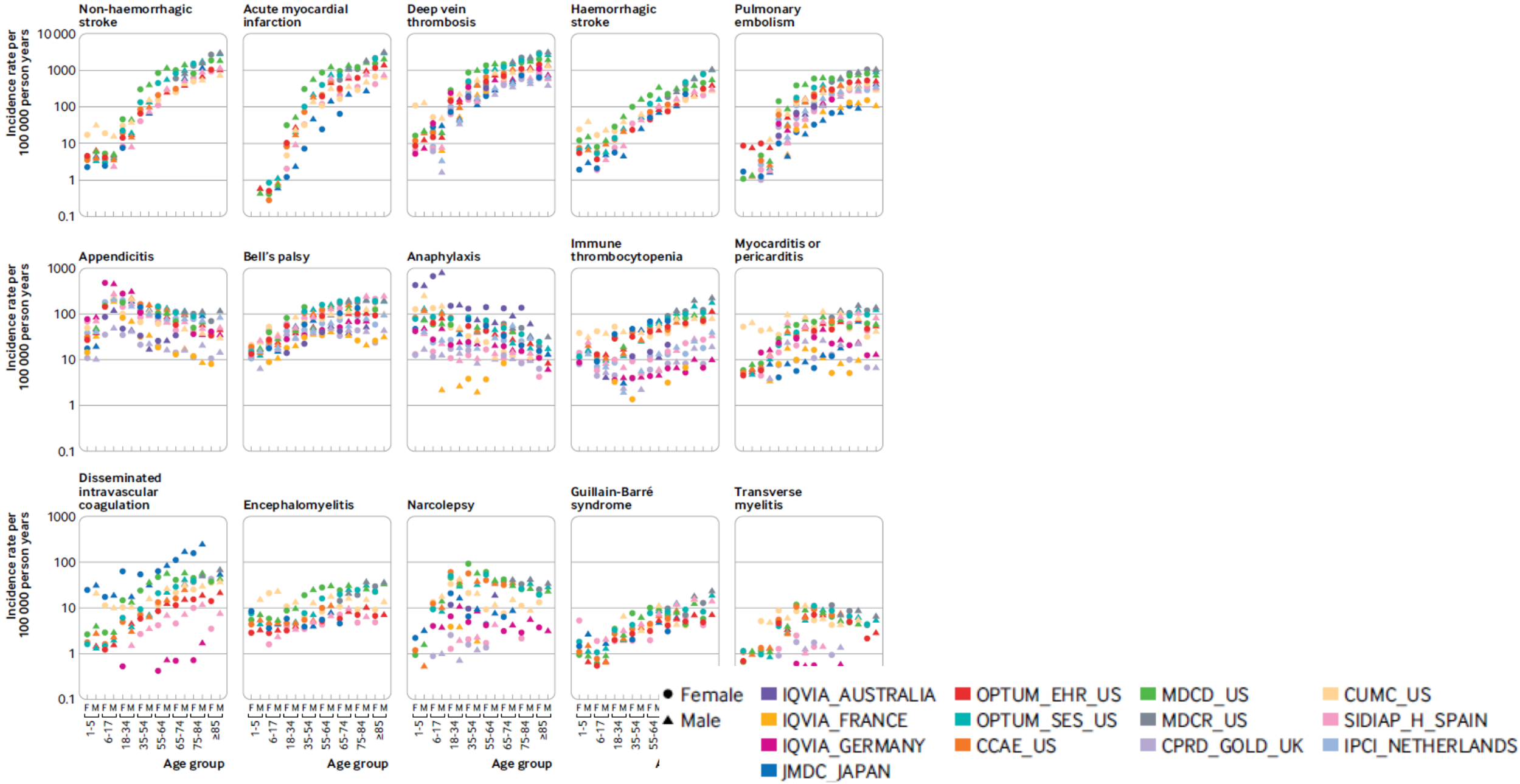
Events of interests were 15 prespecified AESIs (non-haemorrhagic and haemorrhagic stroke, acute myocardial infarction, deep vein thrombosis, pulmonary embolism, anaphylaxis, Bell's palsy, myocarditis or pericarditis, narcolepsy, appendicitis, immune thrombocytopenia, disseminated intravascular coagulation, encephalomyelitis (including acute disseminated encephalomyelitis), Guillain-Barré syndrome, and transverse myelitis). Incidence rates of AESIs were stratified by age, sex, and database. Rates were pooled across databases using random effects meta-analyses and classified according to the frequency categories of the Council for International Organizations of Medical Sciences.

RESULTS

Background rates varied greatly between databases. Deep vein thrombosis ranged from 387 (95% confidence interval 370 to 404) per 100 000 person years in UK CPRD GOLD data to 1443 (1416 to 1470) per 100 000 person years in US IBM MarketScan Multi-State Medicaid data among women aged 65 to 74 years. Some AESIs increased with age. For example, myocardial infarction rates in men increased from 28 (27 to 29) per 100 000 person years among those aged 18-34 years to 1400 (1374 to 1427) per 100 000 person years in those older than 85 years in US Optum electronic health record data. Other AESIs were more common in young people. For example, rates of anaphylaxis among boys and men were 78 (75 to 80) per 100 000 person years in those aged 6-17 years and 8 (6 to 10) per 100 000 person years in those older than 85 years in Optum electronic health record data. Meta-analytic estimates of AESI rates were classified according to age and sex.

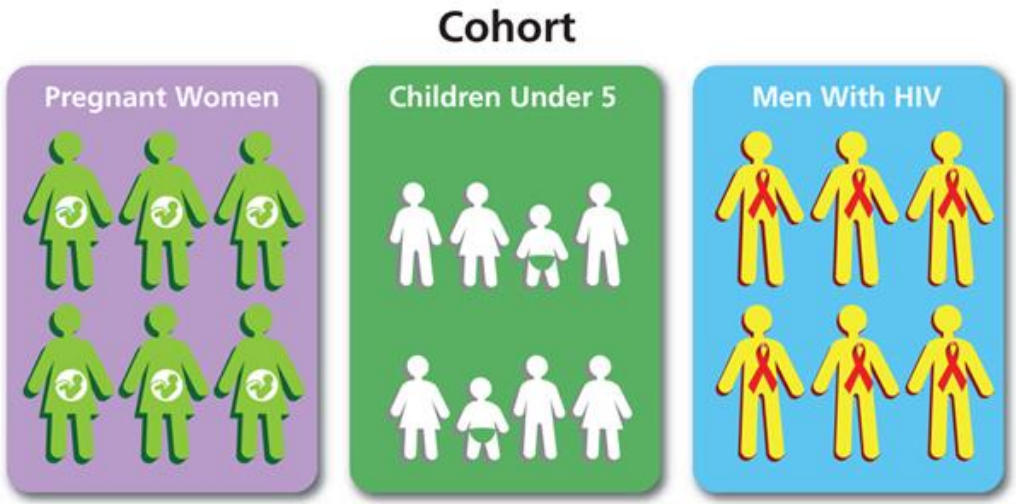
CONCLUSION

This study found large variations in the observed rates of AESIs by age group and sex, showing the need for stratification or standardisation before using


Non-haemorrhagic stroke
Acute myocardial infarction
Deep vein thrombosis
Haemorrhagic stroke
Pulmonary embolism
Appendicitis
Bell's palsy
Anaphylaxis
Immune thrombocytopenia
Myocarditis or pericarditis
Disseminated intravascular coagulation
Encephalomyelitis
Narcolepsy
Guillain-Barré syndrome
Transverse myelitis
Incidence rate per 100 000 person years
10 000
1000
100
10
1
0.1
F M
1-5
6-17
18-34
35-54
55-64
65-74
75-84
≥85
Age group
Female
Male
IQVIA_AUSTRALIA
IQVIA_FRANCE
IQVIA_GERMANY
JMDC_JAPAN
OPTUM_EHR_US
OPTUM_SES_US
CCAE_US
MDCD_US
MDCR_US
CPRD_GOLD_UK
CUMC_US
SIDIAP_H_SPAIN
IPCI_NETHERLANDS

‘Cohort’

- 로마군에서 대대를 가리키는 cohors
- 코호트- 특정 기간 동안 공통된 특성이나 경험을 갖는 사용자 집단
  - Framingham Heart Study
  - 강화 코호트



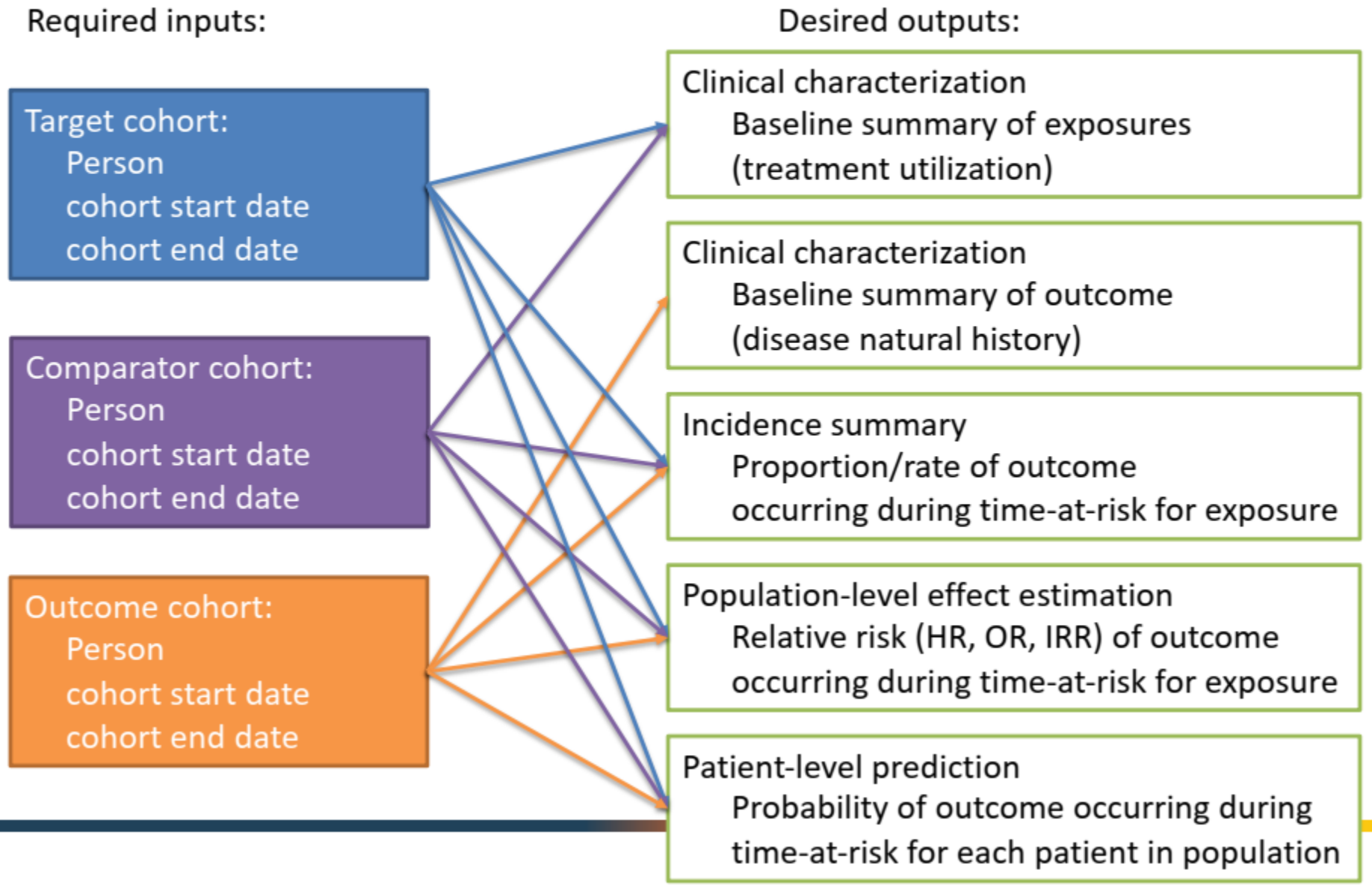
Definition of Cohort in OHDSI

- A cohort = a set of persons who satisfy one or more inclusion criteria for a duration of time.

Features of “OHDSI-defined Cohort”

- One person may belong to multiple cohorts
- One person may belong to the same cohort for multiple different time periods
- One person may not belong to the same cohort multiple times during the same period of time
- A cohort may have zero or more members

The common building block of all observational analysis: cohorts



코호트 연구의 대상자 정의

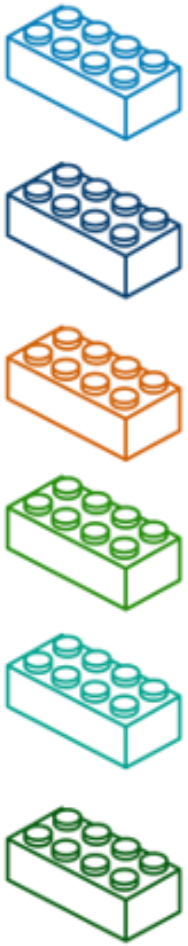
- Prevalent user
  - 한번이라도 약물 처방 있는 대상자
  - 약물 지속 사용환자
    - 약물요법에 잘 적응, 치료효과가 있을 가능성이 큼
  - 대상자마다 약물사용기간 다름

- Incident user (New user)
  - wash-out period 적용
    (예. 이전 1년 이내 약물사용 없었던 사람)
  - 대규모 인구집단 자료, 장기간의 자료 필요

Questions to answer when defining a cohort

- What initial event(s) define cohort entry?
- What inclusion criteria are applied to the initial events?
- What defines a person's cohort exit?

Defining cohort components



Conditions

Drugs

Procedures

Measurements

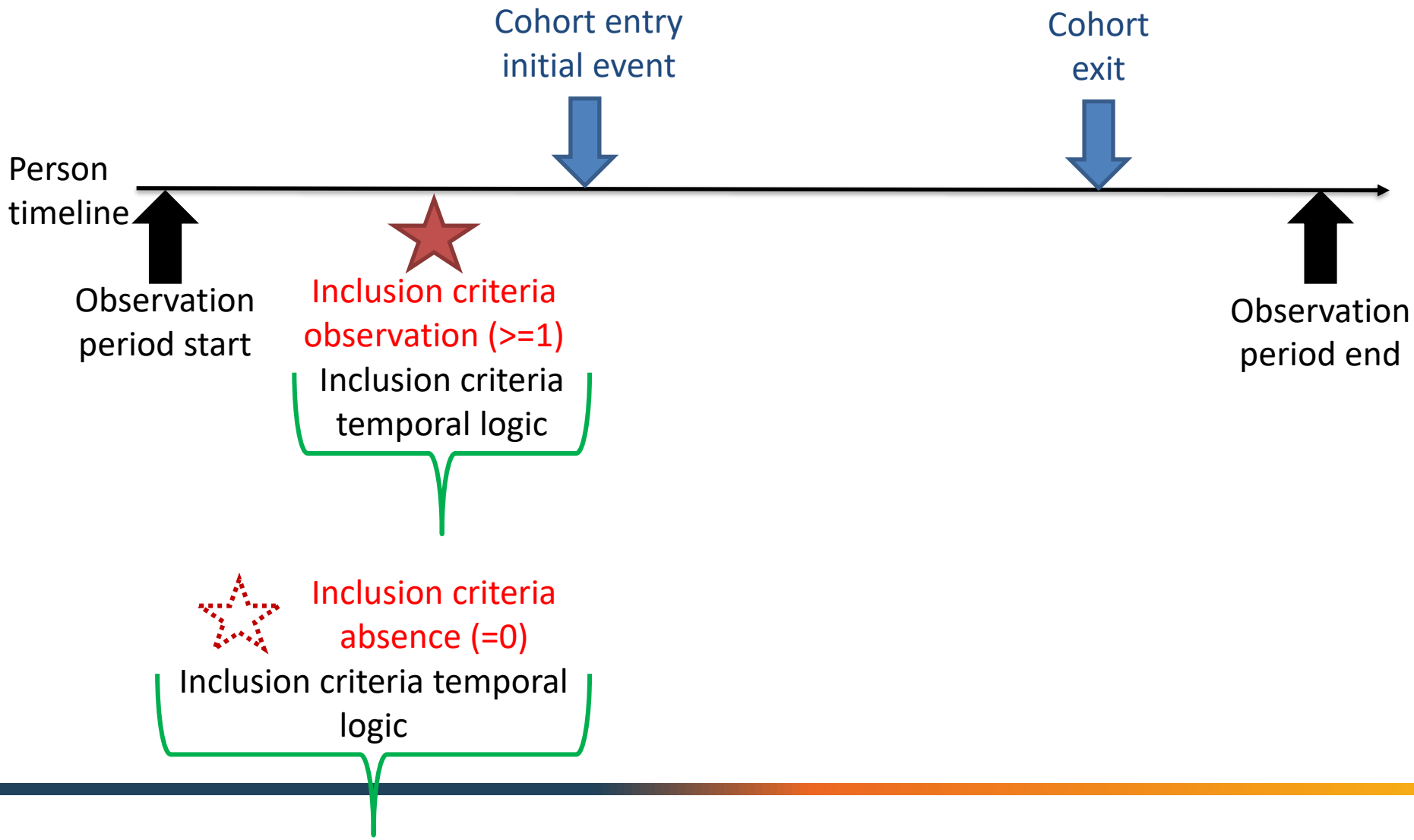
Observations

Visits

Defining cohort components

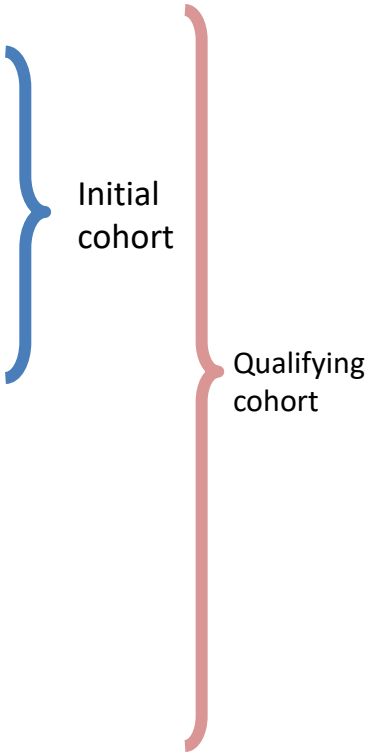
- Domain: A Domain defines the set of allowable Concepts for the standardized fields in the CDM tables.
  - Ex: Condition, Drug, Procedure, Measurement
- Conceptset: An expression that defines one or more concepts encompassing a clinical entity of interest
  - Ex: Concepts for T2DM, concepts for antidiabetic drugs
- Domain-specific attribute:
  - Ex: DRUG_EXPOSURE: Days supply; MEASUREMENT: value_as_number, high_range
- Temporal logic: the time intervals within which the relationship between an inclusion criteria and an event is evaluated
  - Ex: Indicated condition must occur during 365d prior to or on exposure start

Dissecting the anatomy of a cohort definition



Define cohort: Process flow for formally defining a cohort in ATLAS

- **Cohort entry criteria**
  - **Initial events**
    - 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 the same value (such as procedures or measurements).
  - **Initial event inclusion criteria**
  - **Additional qualifying inclusion criteria**
    - The qualifying cohort will be defined as all persons who have an initial event, satisfy the initial event inclusion criteria, and fulfill all additional qualifying inclusion criteria
    - Each qualifying inclusion criteria will be evaluated to determine the impact of the criteria on the attrition of persons from the initial cohort
- **Cohort exit criteria**

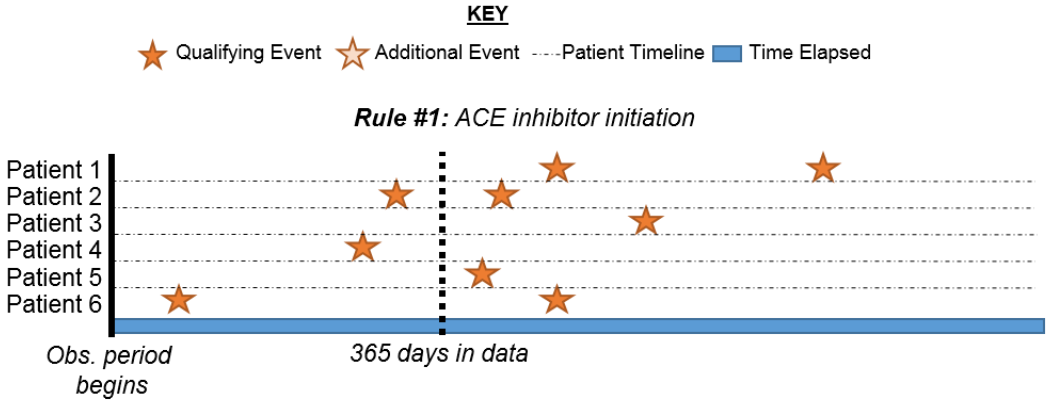


Define cohort: Process flow for formally defining a cohort in ATLAS

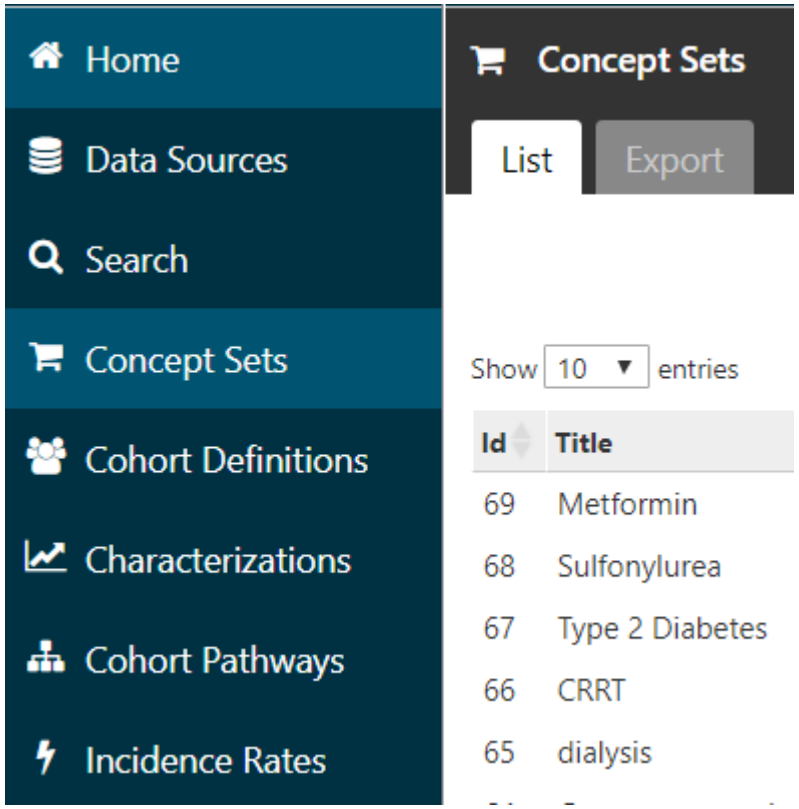
Initial cohort

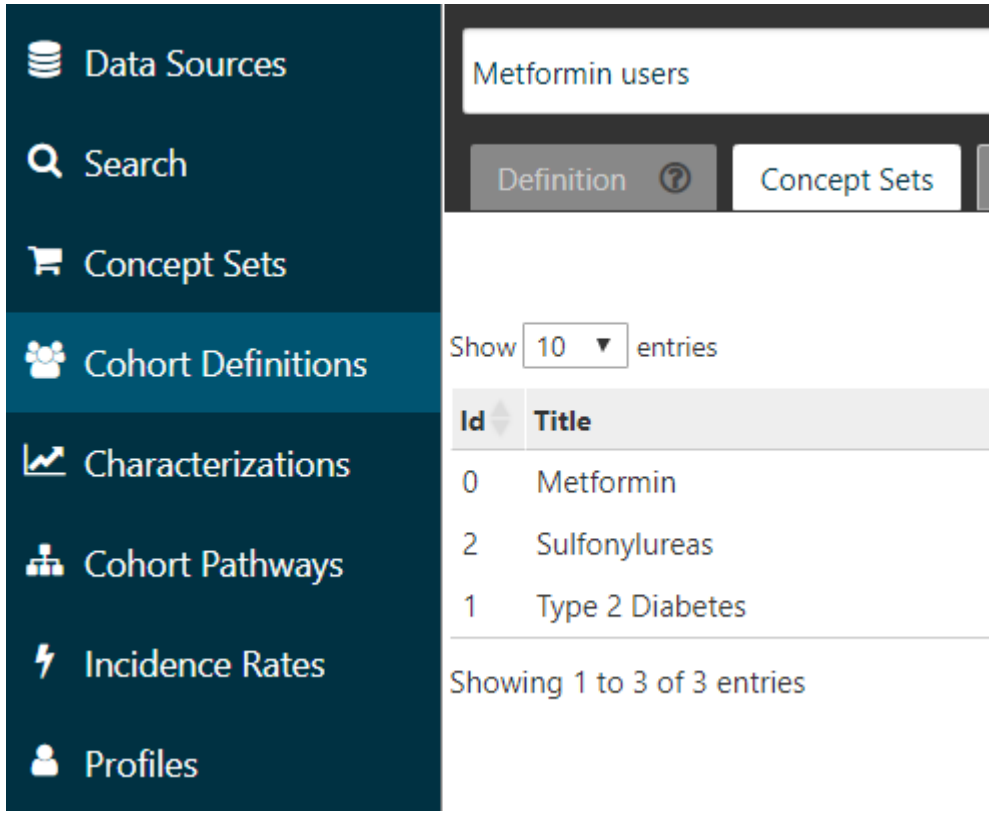
Qualifying cohort

Explaining patient eligibility by criteria applied



Concept Sets vs Concept Sets





Temporal logic: before vs after

Restrict intial events to:

having all ▼ of the following criteria:

with at least ▼ 1 ▼ using all occurrences of:

a condition occurrence of Type 2 Diabetes ▼

where event starts between All ▼ days Before ▼ and 0 ▼ days Before ▼ ndex start date *add additional constraint*

☐ restrict to the same visit occurrence

☐ allow events from outside observation period

Restrict intial events to:

having all ▼ of the following criteria:

with at least ▼ 1 ▼ using all occurrences of:

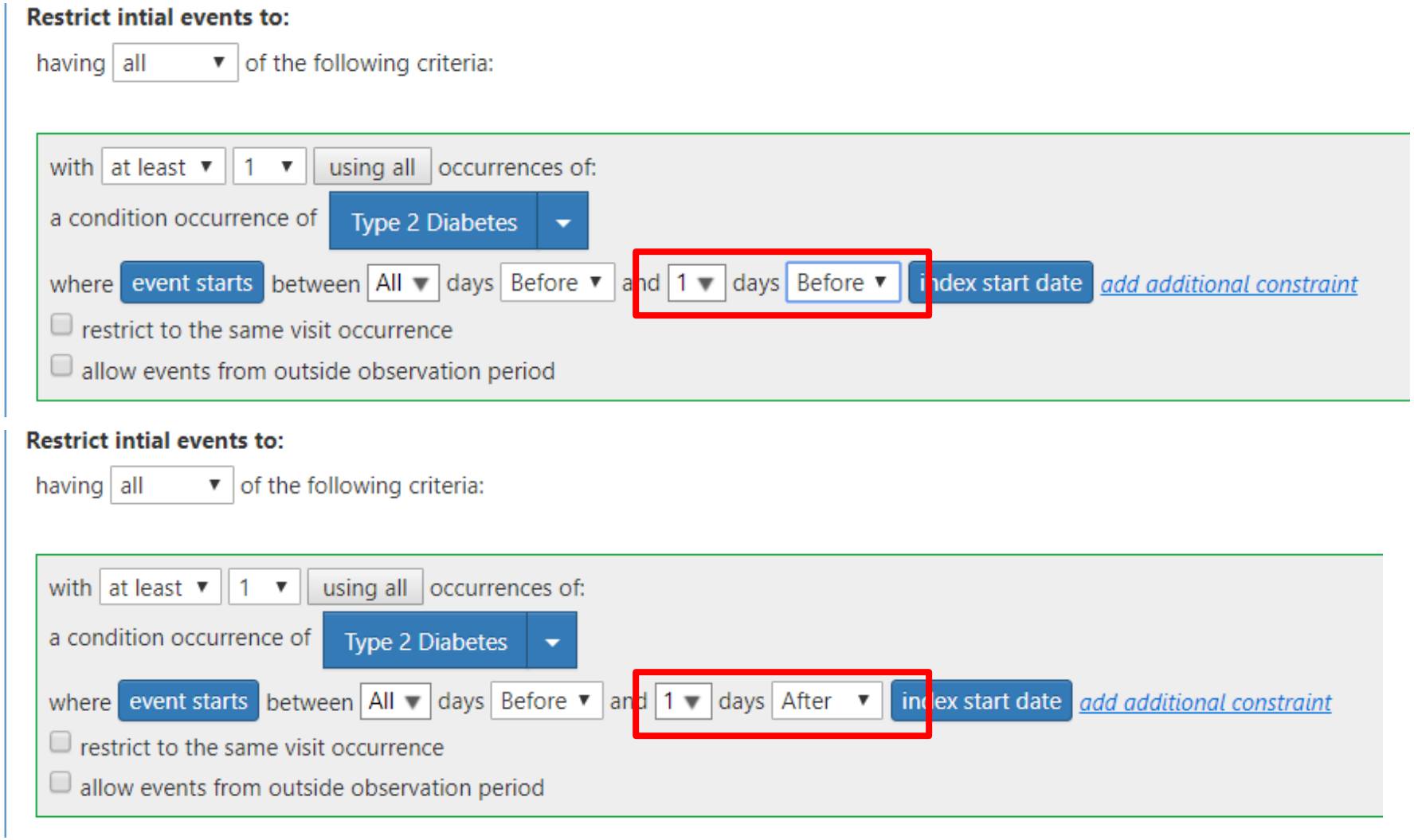
a condition occurrence of Type 2 Diabetes ▼

where event starts between All ▼ days Before ▼ and 0 ▼ days After ▼ ndex start date *add additional constraint*

☐ restrict to the same visit occurrence

☐ allow events from outside observation period

Temporal logic: before vs after

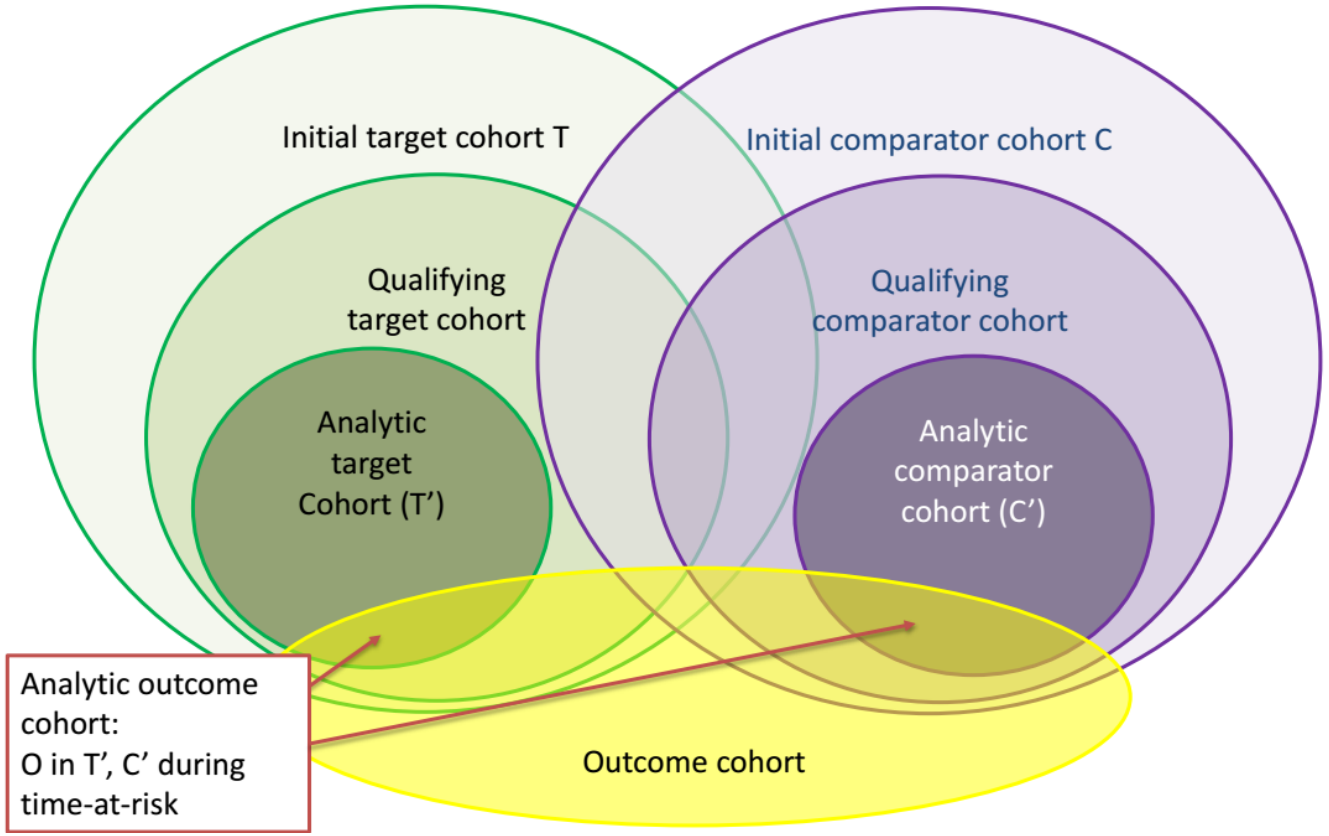


Temporal logic: before vs after

```
 98 FROM
 99 (
100   SELECT co.*
101   FROM @cdm_database_schema.CONDITION_OCCURRENCE co
102   JOIN #Codesets codesets on ((co.condition_concept_id = codesets.concept_id and codesets.codeset_id =
    1))
103 ) C
104
105
106 -- End Condition Occurrence Criteria
107
    ) A on A.person_id = P.person_id  AND A.START_DATE >= P.OP_START_DATE AND A.START_DATE <= P.OP_END_DATE A
108 ND A.START_DATE >= P.OP_START_DATE AND A.START_DATE <=
    DATEADD(day,-1,P.START_DATE)
109 GROUP BY p.person_id, p.event_id
110 HAVING COUNT(A.TARGET_CONCEPT_ID) >= 1
111 -- End Correlated Criteria
112
```

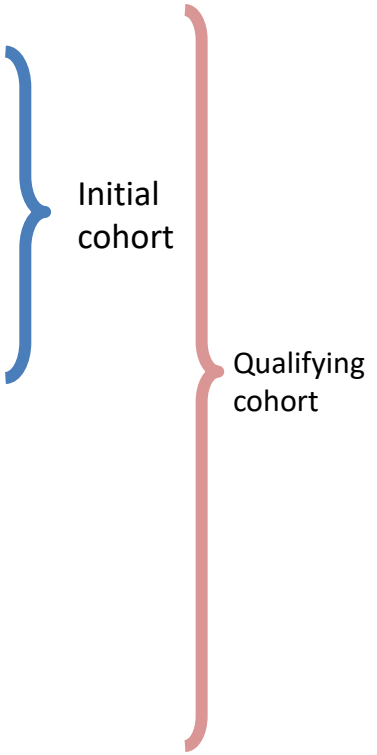
```
 98 FROM
 99 (
100   SELECT co.*
101   FROM @cdm_database_schema.CONDITION_OCCURRENCE co
102   JOIN #Codesets codesets on ((co.condition_concept_id = codesets.concept_id and codesets.codeset_id =
    1))
103 ) C
104
105
106 -- End Condition Occurrence Criteria
107
    ) A on A.person_id = P.person_id  AND A.START_DATE >= P.OP_START_DATE AND A.START_DATE <= P.OP_END_DATE A
108 ND A.START_DATE >= P.OP_START_DATE AND A.START_DATE <=
    DATEADD(day,1,P.START_DATE)
109 GROUP BY p.person_id, p.event_id
110 HAVING COUNT(A.TARGET_CONCEPT_ID) >= 1
111 -- End Correlated Criteria
112
113   ) CQ on E.person_id = CQ.person_id and E.event_id = CQ.event_id
114   GROUP BY E.person_id, E.event_id
115   HAVING COUNT(index_id) = 1
116 ) G
117 -- End Criteria Group
118 ) AC on AC.person_id = pe.person_id and AC.event_id = pe.event_id
119
120 ) QE
121 WHERE QE.ordinal = 1
122 ;
```

Define cohort: A database is full of cohorts, some of which may represent valid comparisons

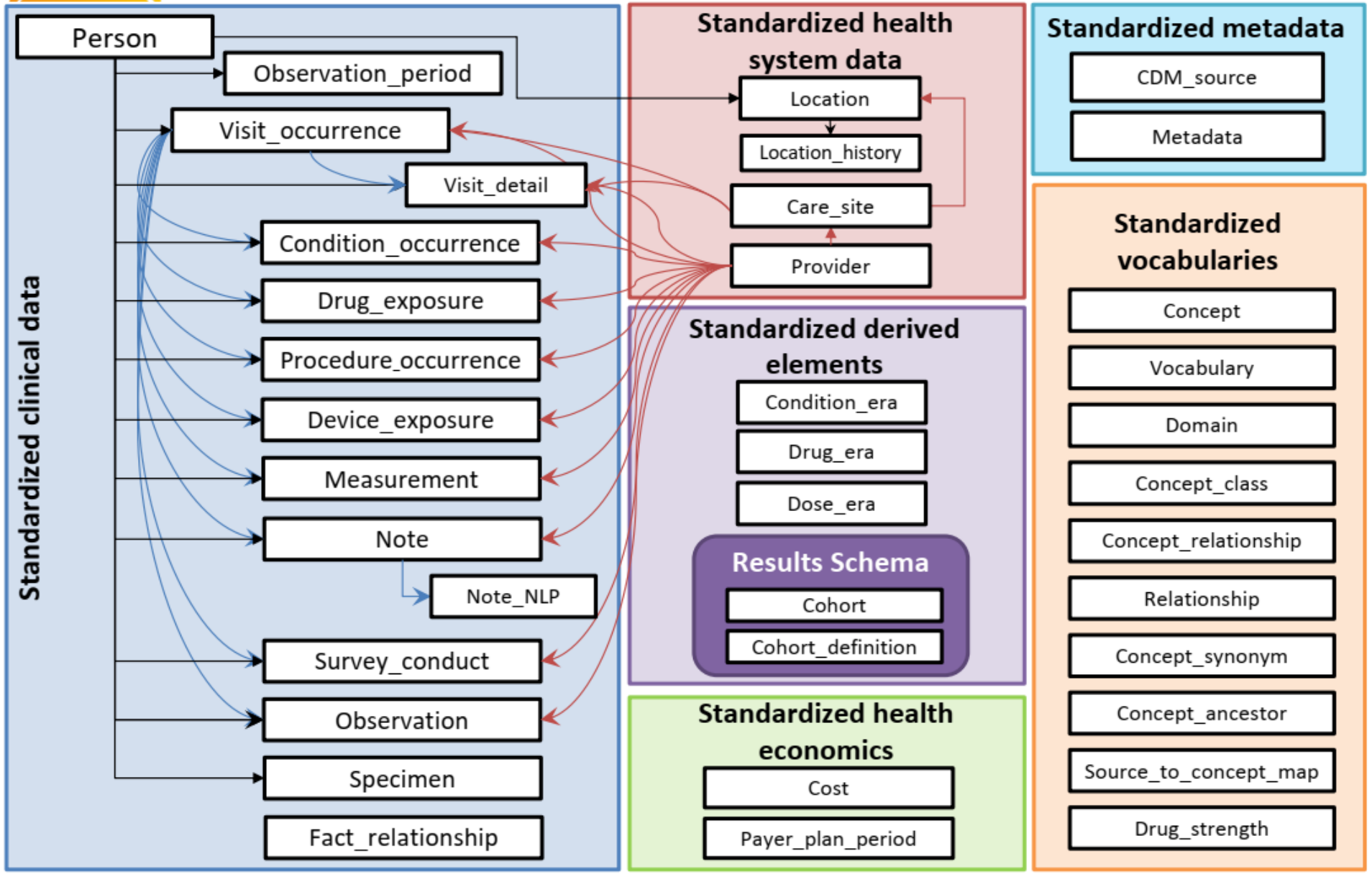


Define cohort: Process flow for formally defining a cohort in ATLAS

- **Cohort entry criteria**
  - **Initial events**
    - 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 the same value (such as procedures or measurements).
  - **Initial event inclusion criteria**
  - **Additional qualifying inclusion criteria**
    - The qualifying cohort will be defined as all persons who have an initial event, satisfy the initial event inclusion criteria, and fulfill all additional qualifying inclusion criteria
    - Each qualifying inclusion criteria will be evaluated to determine the impact of the criteria on the attrition of persons from the initial cohort
- **Cohort exit criteria**



CDM Version 6 Key Domains



OMOP CDM Principles

- Patient centric
- Vocabulary and Data Model are blended
- Domain-oriented concepts
- Accommodates data from various sources
- Preserves data provenance
- Extendable & Evolving
- Database Platform Independent

OMOP CDM Standard Domain Features

Feature	Description & Purpose	Field Name Convention	Example
Patient centric	Every domain table has **patient identifier**. Patient data can be retrieved independently from other domains.	**person_id**	person_id 123
Unique domain identifiers	Ever domain table has a unique primary key to identify domain entities.	<entity>**_id**	condition_occurrence_id 470985
Standard concept from a respective vocabulary domain	Integration with the Vocabulary. Foreign key into the Standard Vocabulary for **Standard Concept.**	<entity>_concept_id	condition_concept_id 313217 (SNOMED "Atrial Fibrillation")
Source value	Provenance. Verbatim information from the source data, **not to be used** by any standard analytics.	<entity>**_source_value**	condition_source_value 427.31 (ICD9CM "Atrial Fibrillation")
Source concept from a respective vocabulary domain	Provenance. Foreign key into Standard Vocabulary for **Source Concept**.	<entity>**_source_concept_id**	condition_source_concept_id 44821957 (ICD9CM "Atrial Fibrillation")
Source type	Provenance. Foreign key into Vocabulary for the **origin of the data.**	<entity>**_type_concept_id**	condition_type_concept_id 38000199 ("Inpatient header – primary")

OMOP Common Vocabulary Model

What it is

- **Standardized structure** to house existing vocabularies used in the public domain
- **Compiled standards** from disparate public and private sources and some OMOP-grown concepts

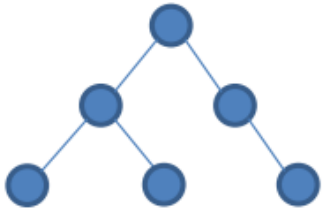
What it's not

- **Static dataset** – the vocabulary updates regularly to keep up with the continual evolution of the sources
- **Finished product** – vocabulary maintenance and improvement is ongoing activity that requires community participation and support

Structure of OMOP Vocabulary

All content: concepts in **`concept`**

Direct relationships between concepts in **`concept_relationship`**

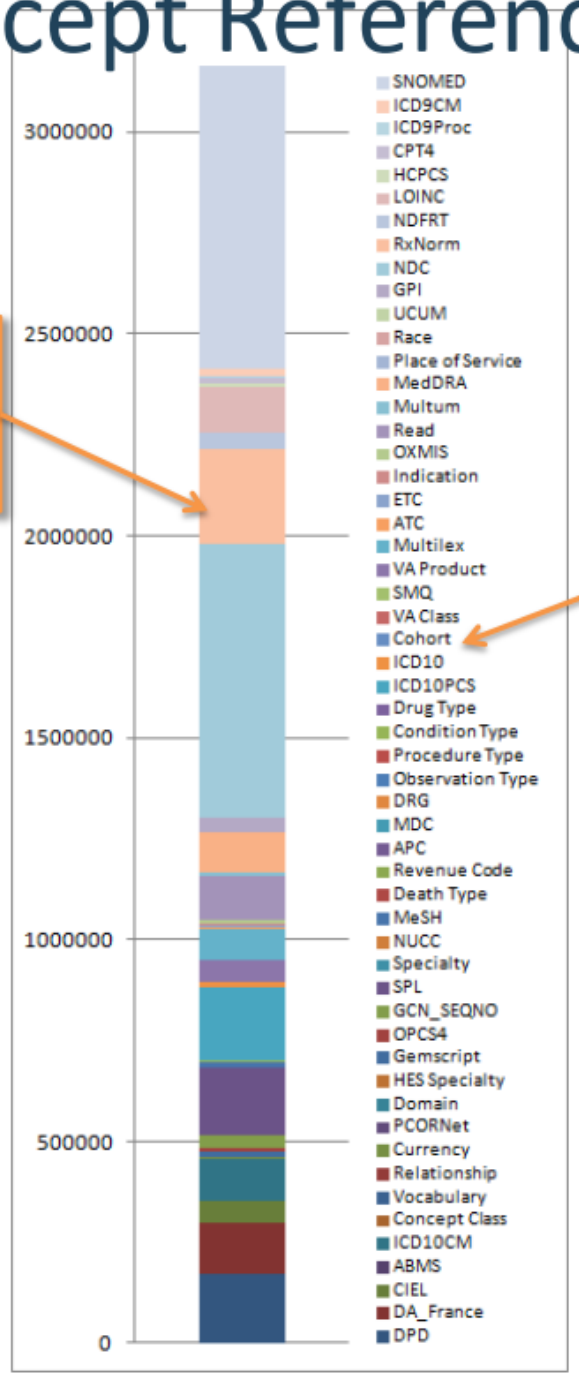


Multi-step hierarchical relationships pre-processed into **`concept_ancestor`**

Single Concept Reference Table

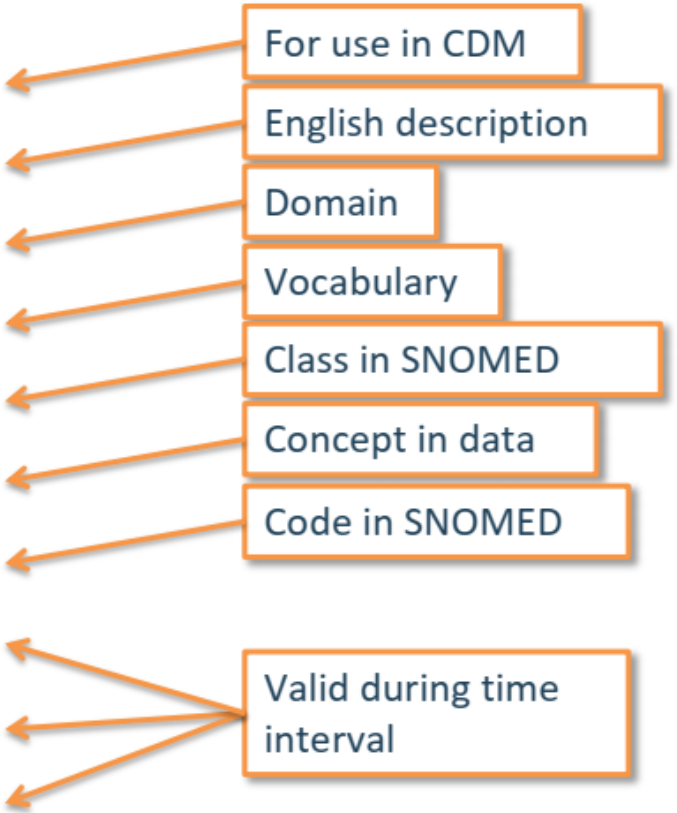
All vocabularies stacked up in one table

Vocabulary ID

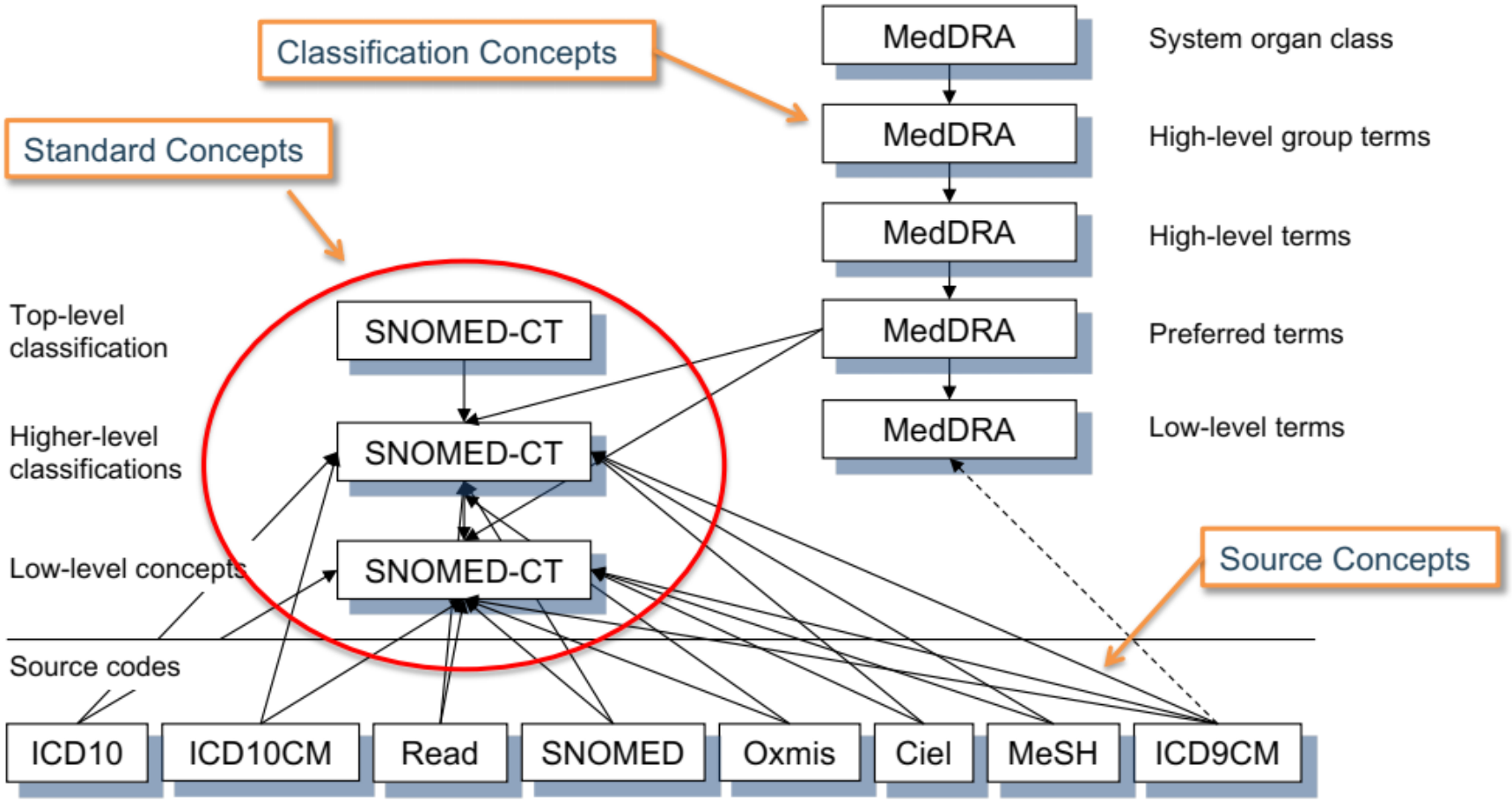


What's in a Concept

CONCEPT_ID	313217
CONCEPT_NAME	Atrial fibrillation
DOMAIN_ID	Condition
VOCABULARY_ID	SNOMED
CONCEPT_CLASS_ID	Clinical Finding
STANDARD_CONCEPT	S
CONCEPT_CODE	49436004
VALID_START_DATE	01-Jan-1970
VALID_END_DATE	31-Dec-2099
INVALID_REASON	



Condition Concepts



Finding the Right Concept #1

1. ..if I know the **ID**

```
SELECT * FROM concept WHERE concept_id = 313217
```

CONCEPT_ID	CONCEPT_ NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_CLASS_ID	STANDARD_CONCEPT	CONCEPT_CODE	VALID_START_DATE	VALID_END_DATE	INVALID_REASON
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	01-Jan-1970	31-Dec-2099	

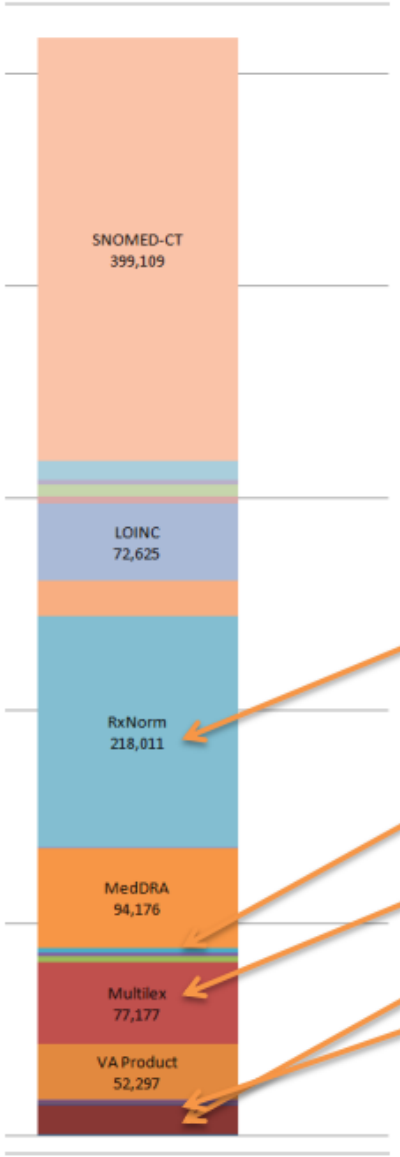
SNOMED code

2. ..if I know the **code**

```
SELECT * FROM concept WHERE concept_code = '49436004'
```

CONCEPT_ID	CONCEPT_ NAME	DOMAIN_ID	VOCABULARY_ID	CONCEPT_CLASS_ID	STANDARD_CONCEPT	CONCEPT_CODE	VALID_START_DATE	VALID_END_DATE	INVALID_REASON
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	01-Jan-1970	31-Dec-2099	

Concept ID versus Concept Code




```
SELECT *
FROM concept
WHERE concept_code = '1001';
```

Same code

Concept_Name	Concept Class	Vocabulary_ID	Concept_Code
Antipyrine	Ingredient	RxNorm	1001
Aceprometazine maleate	Ingredient	BDPM	1001
Serum	Specimen	CIEL	1001
methixene hydrochloride	Ingredient	Multilex	1001
Brompheniramine Maleate, 10 mg/mL injectable solution	Multum	Multum	1001
ABBOTT COLD SORE BALM 4%/0.06% W/	Drug Product	LPD_Australia	1001
Residential Treatment - Psychiatric	Revenue Code	Revenue Code	1001

Finding the Right Concept #2

3. ..if I know the **name**

```
SELECT * FROM concept WHERE concept_name = 'Atrial fibrillation';
```

CONCEPT _ID	CONCEPT_ NAME	DOMAIN _ID	VOCABULARY _ID	CONCEPT_ CLASS_ID	STANDARD_ CONCEPT	CONCEPT_ CODE
313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004
44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code		427.31
35204953	Atrial fibrillation	Condition	MedDRA	PT	C	10003658
45500085	Atrial fibrillation	Condition	Read	Read		G573000
45883018	Atrial fibrillation	Meas Value	LOINC	Answer	S	LA17084-7

Finding the Right Concept #3

1. if don't know any of this, but I know the code in another vocabulary

ICD-9 is not a Standard Concept

```
SELECT * FROM concept WHERE concept_code = '427.31';
```

CONCEPT _ID	CONCEPT_ NAME	DOMAIN _ID	VOCABULARY _ID	CONCEPT_ CLASS_ID	STANDARD_ CONCEPT	CONCEPT_ CODE
44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code		427.31

```
SELECT * FROM concept_relationship WHERE concept_id_1 = 44821957;
```

Mapping to different vocabularies

Kind of relationship

_ID_1	CONCEPT ID_2	RELATIONSHIP _ID	VALID_START _DATE	VALID_END _DATE	INVALID _REASON
44821957	21001551	ICD9CM - FDB Ind	01-Oct-13	31-Dec-2099	
44821957	35204953	ICD9CM - MedDRA	01-Jan-70	31-Dec-2099	
44821957	44824248	Is a	01-Oct-14	31-Dec-2099	
44821957	44834731	Is a	01-Oct-14	31-Dec-2099	
44821957	313217	Maps to	01-Jan-70	31-Dec-2099	

Mapping = Translating

Step 1. Lookup the Source Concept

```
SELECT * FROM concept WHERE concept_code = '427.31';
```

CONCEPT _ID	CONCEPT_ NAME	DOMAIN _ID	VOCABULARY _ID	CONCEPT_ CLASS_ID	STANDARD_ CONCEPT	CONCEPT_ CODE
44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code		427.31

Step 2. Translate to Standard

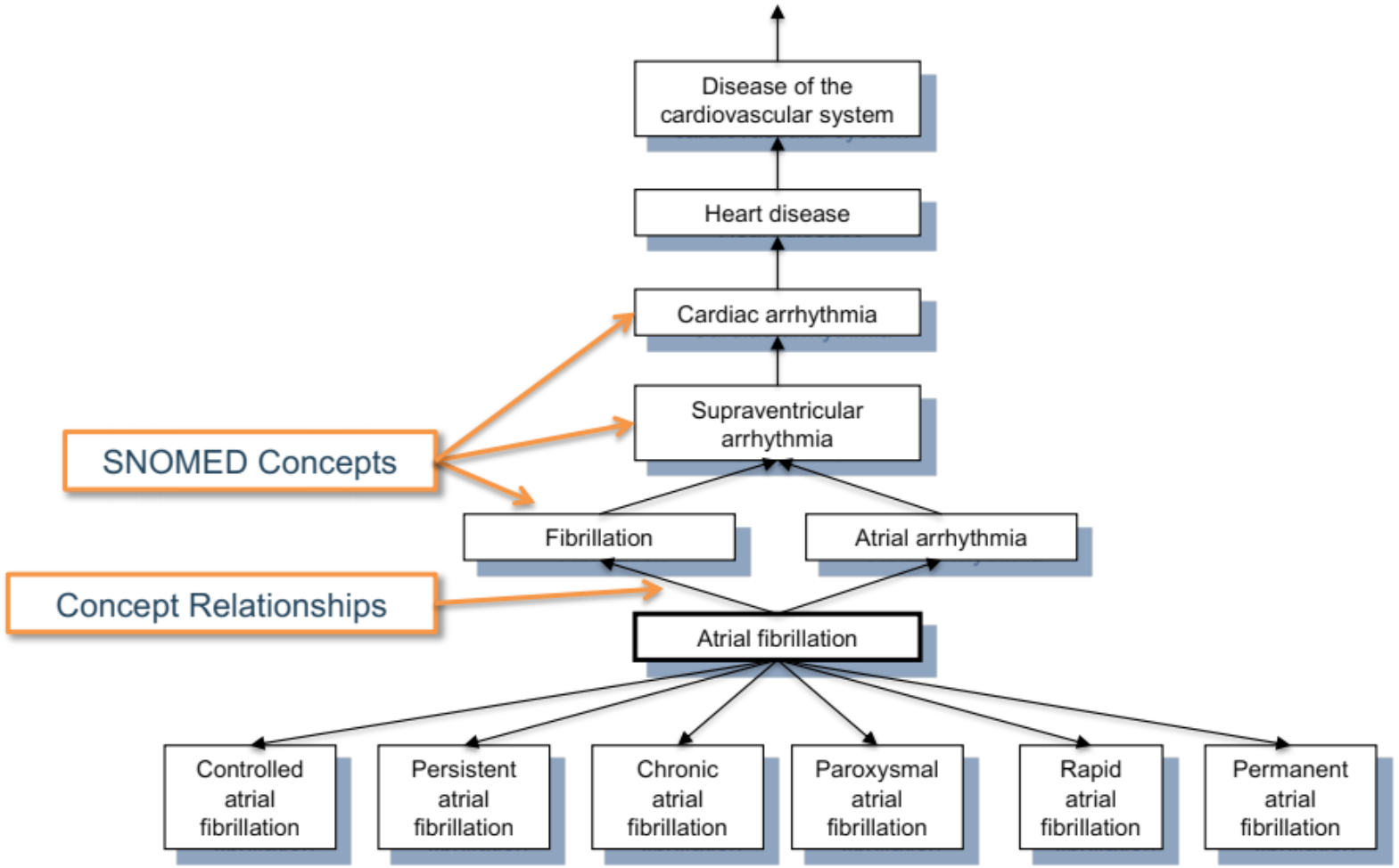
```
SELECT * FROM concept_relationship WHERE concept_id_1 = 44821957
    AND relationship_id = 'Maps to';
```

CONCEPT _ID_1	CONCEPT _ID_2	RELATIONSHIP _ID	VALID_START _DATE	VALID_END _DATE	INVALID _REASON
44821957	313217	Maps to	01-Jan-1970	31-Dec-2099	

Step 3. Check out the translated Concept

```
SELECT * FROM concept WHERE concept_id = 313217;
```

Reason #2: Disease Hierarchy



Exploring Relationships

```
SELECT *
FROM concept_relationship
WHERE concept_id_1 = 313217
```

CONCEPT ID_1	CONCEPT ID_2	RELATIONSHIP _ID
313217	4232697	Subsumes
313217	4181800	Focus of
313217	35204953	SNOMED - MedDRA eq
313217	4203375	Asso finding of
313217	4141360	Subsumes
313217	4119601	Subsumes
313217	4117112	Subsumes
313217	4232691	Subsumes
313217	4139517	Due to of
313217	4194288	Asso finding of
313217	44782442	Subsumes
313217	44783731	Focus of
313217	21003018	SNOMED - ind/CI
313217	40248987	SNOMED - ind/CI
313217	21001551	SNOMED - ind/CI
313217	21001540	SNOMED - ind/CI
313217	45576876	Mapped from
313217	44807374	Asso finding of
313217	21013834	SNOMED - ind/CI
313217	21001572	SNOMED - ind/CI
313217	21001606	SNOMED - ind/CI
313217	21003176	SNOMED - ind/CI
313217	4226399	Is a
313217	500001801	SNOMED - HOI
313217	500002401	SNOMED - HOI
313217	4119602	Subsumes
313217	40631039	Subsumes
313217	4108832	Subsumes
313217	21013671	SNOMED - ind/CI
313217	21013390	SNOMED - ind/CI
313217	313217	Maps to
313217	44821957	Mapped from
313217	2617597	Mapped from
313217	45500085	Mapped from
313217	313217	Mapped from
313217	45951191	Mapped from
313217	21013856	SNOMED - ind/CI
313217	21001575	SNOMED - ind/CI
313217	21001594	SNOMED - ind/CI

Related Concepts

Relationship ID

Exploring Relationships

```
SELECT cr.relationship_id, c.*
FROM concept_relationship cr
JOIN concept c ON cr.concept_id_2 = c.concept_id
WHERE cr.concept_id_1 = 313217
```

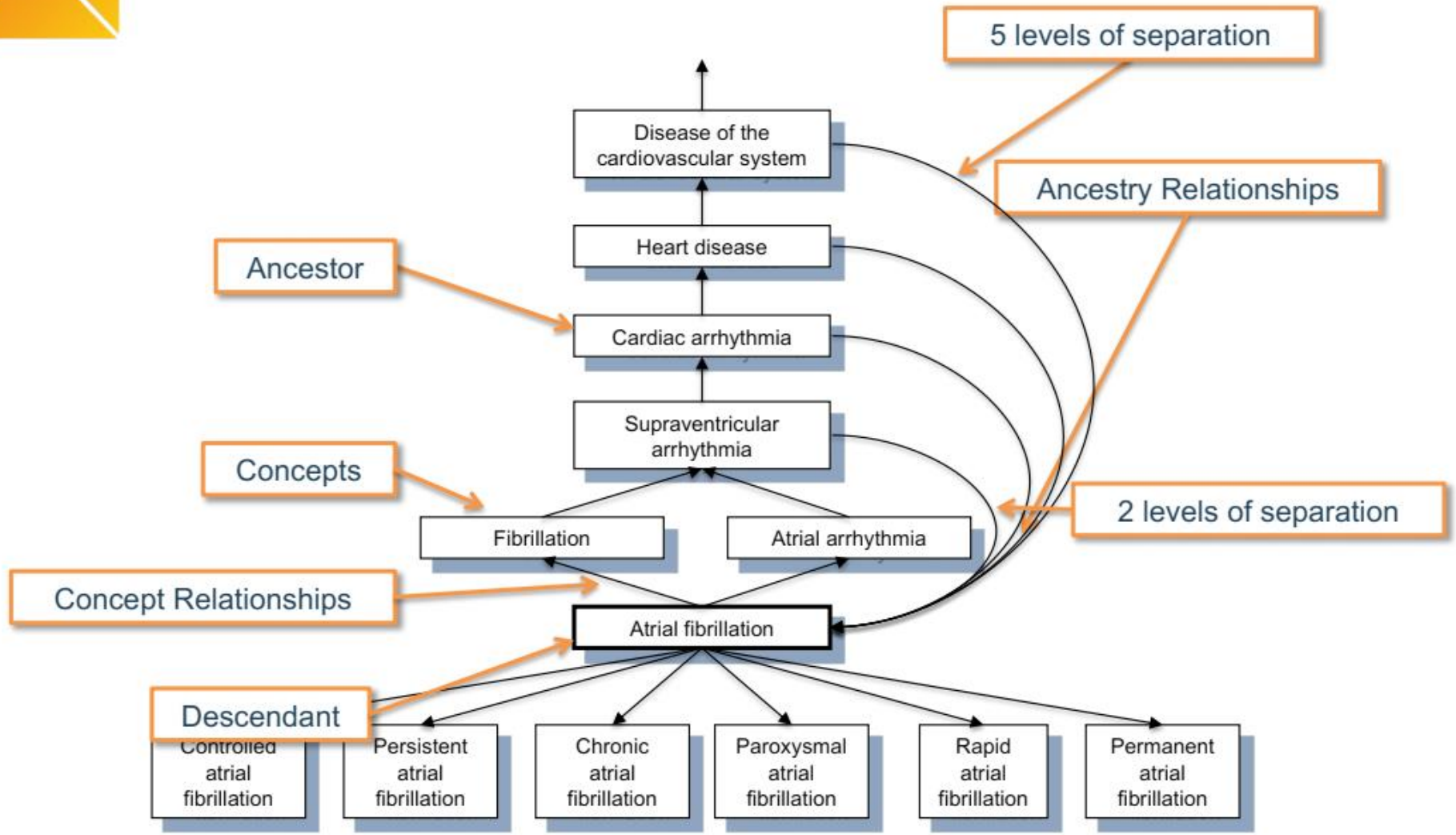
Find out related concept

Ancestor concepts

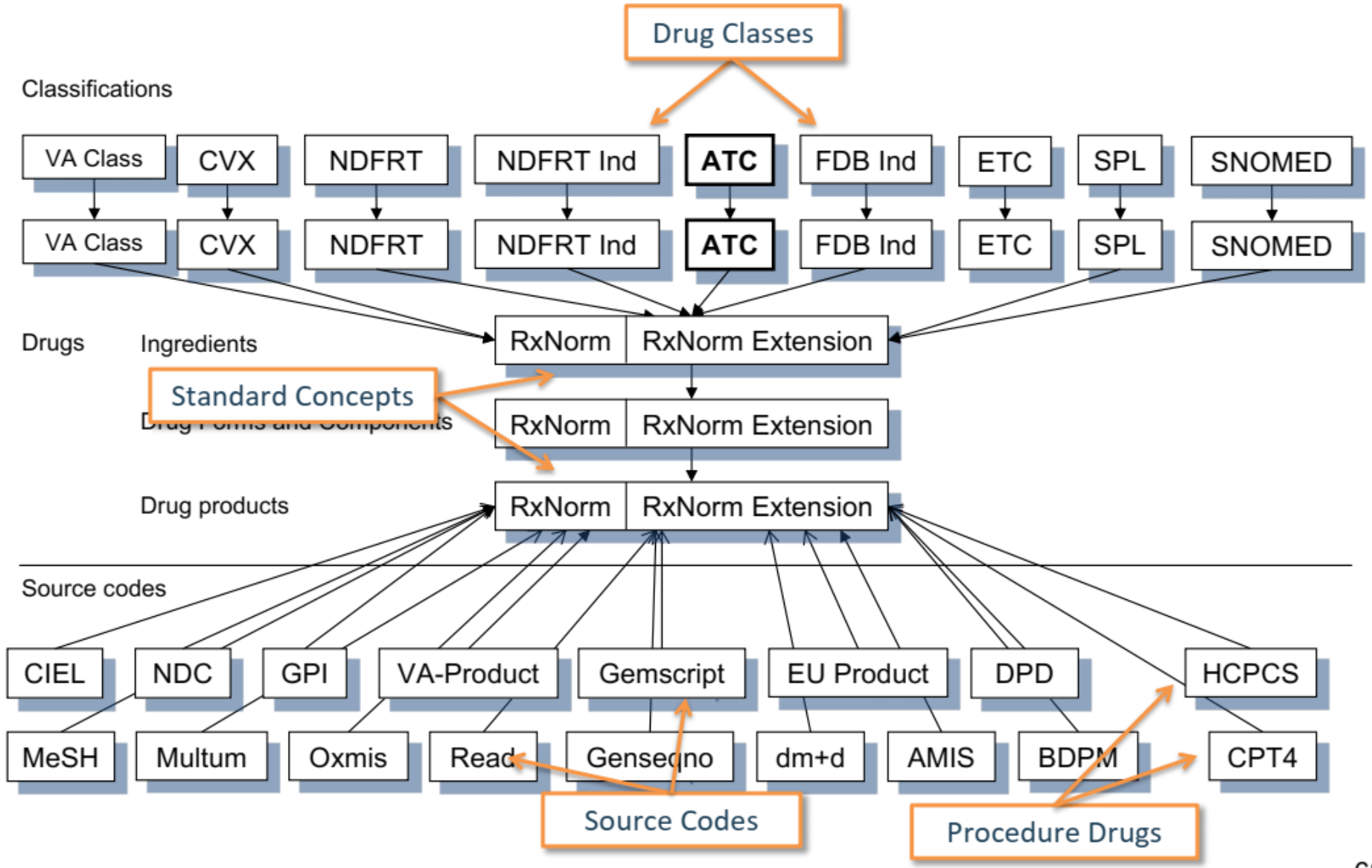
relationship_id	concept_id	concept_name	domain_id	vocabulary_id	concept_class_id	standard_concept	concept_code	valid_start_date	valid_end_date	invalid_reason
Asso finding of	4194288		Observation	SNOMED	Context-dependent	S	312442005	1/1/1970 0:00	12/31/2099 0:00	NULL
Asso finding of	4203375		Observation	SNOMED	Context-dependent	S	433276002	1/31/2009 0:00	12/31/2099 0:00	NULL
Asso finding of	42689685	Atrial fibrillation not d[illegible]sted	Observation	SNOMED	Context-dependent	S	1.06706E+15	4/1/2017 0:00	12/31/2099 0:00	NULL
Asso finding of	44807374	Atrial fibrillation excluded	Observation	SNOMED	Context-dependent	S	8.16401E+14	4/1/2014 0:00	12/31/2099 0:00	NULL
Concept poss_eq from	40323929	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	155364009	1/1/1970 0:00	3/11/2016 0:00	U
Concept poss_eq from	40345197	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	266306001	1/1/1970 0:00	3/11/2016 0:00	U
Due to of	4139517	Transient cerebral ischemia due to atrial fibrillation	Condition	SNOMED	Clinical Finding	S	426814001	1/1/1970 0:00	12/31/2099 0:00	NULL
Focus of	4209991	Insertion of pacemaker for control of atrial fibrillation	Procedure	SNOMED	Procedure	S	449863006	1/31/2012 0:00	12/31/2099 0:00	NULL
Has finding site	4242112	Atrial structure	Spec Anatomic Site	SNOMED	Body Structure	S	59652004	1/1/1970 0:00	12/31/2099 0:00	NULL
Is a	4226399	Fibrillation	Condition	SNOMED	Clinical Finding	S	40593004	1/1/1970 0:00	12/31/2099 0:00	NULL
Is a	4068155	Atrial arrhythmia	Condition	SNOMED	Clinical Finding	S	17366009	1/1/1970 0:00	12/31/2099 0:00	NULL
Mapped from	40323929	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	155364009	1/1/1970 0:00	3/11/2016 0:00	U
Mapped from	2617597	Patient with heart failure and atrial fibrillation documented to be on warfarin therapy	Observation	HCPCS	HCPCS	NULL	G8183	1/1/1970 0:00	11/11/2014 0:00	D
Mapped from	45576876	Unspecified atrial fibrillation	Condition	ICD10CM	5-char billing code	NULL	I48.91	12/30/2006 0:00	12/31/2099 0:00	NULL
Mapped from	45500085	Atrial fibrillation	Condition	Read	Read	NULL	G573000	1/1/1970 0:00	12/31/2099 0:00	NULL
Mapped from	45611600	Atrial Fibrillation	Condition	MeSH	Main Heading	NULL	D001281	1/1/1970 0:00	12/31/2099 0:00	NULL
Mapped from	40345197	Fibrillation - atrial	Condition	SNOMED	Clinical Finding	NULL	266306001	1/1/1970 0:00	3/11/2016 0:00	U
Mapped from	45951191	Atrial Fibrillation	Condition	CIEL	Diagnosis	NULL	148203	11/3/2007 0:00	12/31/2099 0:00	NULL
Mapped from	313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	1/1/1970 0:00	12/31/2099 0:00	NULL
Mapped from	44821957	Atrial fibrillation	Condition	ICD9CM	5-dig billing code	NULL	427.31	1/1/1970 0:00	12/31/2099 0:00	NULL
Maps to	313217	Atrial fibrillation	Condition	SNOMED	Clinical Finding	S	49436004	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - HOI	500002401	OMOP Atrial Fibrillation 1	Condition	Cohort	Cohort	C	500002401	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - HOI	500001801	OMOP Qt Prolongation/Torsade De Pointes 1	Condition	Cohort	Cohort	C	500001801	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21005673	Prevention of Thromboembolism in Chronic Atrial Fibrillation	Drug	Indication	Indication	C	5673	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21003176	Tachyarrhythmia	Drug	Indication	Indication	C	3176	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21001542	Supraventricular Tachycardia	Drug	Indication	Indication	C	1542	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - ind/CI	21001594	Disease of Cardiovascular System	Drug	Indication	Indication	C	1594	1/1/1970 0:00	12/31/2099 0:00	NULL
SNOMED - MedDRA eq	35204953	Atrial fibrillation	Condition	MedDRA	PT	C	10003658	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4117112	Controlled atrial fibrillation	Condition	SNOMED	Clinical Finding	S	300996004	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4119601	Lone atrial fibrillation	Condition	SNOMED	Clinical Finding	S	233910005	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4232697	Persistent atrial fibrillation	Condition	SNOMED	Clinical Finding	S	440059007	1/31/2009 0:00	12/31/2099 0:00	NULL
Subsumes	4141360	Chronic atrial fibrillation	Condition	SNOMED	Clinical Finding	S	426749004	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	44782442	Atrial fibrillation with rapid ventricular response	Condition	SNOMED	Clinical Finding	S	1.20041E+14	1/31/2014 0:00	12/31/2099 0:00	NULL
Subsumes	4199501	Rapid atrial fibrillation	Condition	SNOMED	Clinical Finding	S	314208002	1/1/1970 0:00	12/31/2099 0:00	NULL
Subsumes	4119602	Non-rheumatic atrial fibrillation	Condition	SNOMED	Clinical Finding	S	233911009	1/1/1970 0:00	12/31/2099 0:00	NULL

Descendant concepts

Ancestry Relationships: Higher-Level Relationships



Drug Hierarchy



Explore the OMOP Vocabulary

http://athena.ohdsi.org/

Person

PERSON

field	required	type
person_id	Yes	INTEGER
gender_concept_id	Yes	INTEGER
year_of_birth	Yes	INTEGER
month_of_birth	No	INTEGER
day_of_birth	No	INTEGER
birth_datetime	No	DATETIME
death_datetime	No	DATETIME
race_concept_id	Yes	INTEGER
ethnicity_concept_id	Yes	INTEGER
location_id	No	INTEGER
provider_id	No	INTEGER
care_site_id	No	INTEGER
person_source_value	No	VARCHAR(50)
gender_source_value	No	VARCHAR(50)
gender_source_concept_id	Yes	INTEGER
race_source_value	No	VARCHAR(50)
race_source_concept_id	Yes	INTEGER
ethnicity_source_value	No	VARCHAR(50)
ethnicity_source_concept_id	Yes	INTEGER

- 환자 기본정보
- Person 당 한 개 record
- 생년월일 중 년도(year_of_birth)는 필수
- death_datetime: v6.0에서 적용

OBSERVATION PERIOD

OBSERVATION_PERIOD

field	required	type
observation_period_id	Yes	INTEGER
person_id	Yes	INTEGER
observation_period_start_date	Yes	DATE
observation_period_end_date	Yes	DATE
period_type_concept_id	Yes	INTEGER

- DB상 환자 관찰 기간 (데이터 보유 기간)
- EMR 데이터의 경우, person 당 1 개 record 생성 (첫 기록 ~ 마지막 기록)
- Observation period 내에 있는 데이터만 분석에 이용

VISIT OCCURRENCE

VISIT_OCCURRENCE

field	required	type
visit_occurrence_id	Yes	INTEGER
person_id	Yes	INTEGER
visit_concept_id	Yes	INTEGER
visit_start_date	No	DATE
visit_start_datetime	Yes	DATETIME
visit_end_date	No	DATE
visit_end_datetime	Yes	DATETIME
visit_type_concept_id	Yes	INTEGER
provider_id	No	INTEGER
care_site_id	No	INTEGER
visit_source_value	No	VARCHAR(50)
visit_source_concept_id	Yes	INTEGER
admitting_source_concept_id	Yes	INTEGER
admitting_source_value	No	VARCHAR(50)
discharge_to_concept_id	Yes	INTEGER
discharge_to_source_value	No	VARCHAR(50)
preceding_visit_occurrence_id	No	INTEGER

- 외래/입원/응급 등 수진정보
- 방문 유형(visit_concept_id)
  - 9201: Inpatient Visit(입원)
  - 9202 : Outpatient Visit(외래)
  - 9203: Emergency Room Visit(응급)
  - ....

CONDITION OCCURRENCE

CONDITION_OCCURRENCE

field	required	type
condition_occurrence_id	Yes	BIGINT
person_id	Yes	BIGINT
condition_concept_id	Yes	INTEGER
condition_start_date	No	DATE
condition_start_datetime	Yes	DATETIME
condition_end_date	No	DATE
condition_end_datetime	No	DATETIME
condition_type_concept_id	Yes	INTEGER
condition_status_concept_id	Yes	INTEGER
stop_reason	No	VARCHAR(20)
provider_id	No	INTEGER
visit_occurrence_id	No	INTEGER
visit_detail_id	No	INTEGER
condition_source_value	No	VARCHAR(50)
condition_source_concept_id	Yes	INTEGER
condition_status_source_value	No	VARCHAR(50)

- 진단(주진단/부진단), 주호소 등
- 사망진단은 v6.0에서 적용
- 표준용어: SNOMED CT, ICD-O

DRUG EXPOSURE

DRUG_EXPOSURE

field	required	type
drug_exposure_id	Yes	BIGINT
person_id	Yes	BIGINT
drug_concept_id	Yes	INTEGER
drug_exposure_start_date	No	DATE
drug_exposure_start_datetime	Yes	DATETIME
drug_exposure_end_date	No	DATE
drug_exposure_end_datetime	No	DATETIME
verbatim_end_date	No	DATE
drug_type_concept_id	Yes	INTEGER
stop_reason	No	VARCHAR(20)
refills	No	INTEGER
quantity	No	FLOAT
days_supply	No	INTEGER
sig	No	VARCHAR(MAX)
route_concept_id	Yes	INTEGER
lot_number	No	VARCHAR(50)
provider_id	No	INTEGER
visit_occurrence_id	No	INTEGER
visit_detail_id	No	INTEGER
drug_source_value	No	VARCHAR(50)
drug_source_concept_id	Yes	INTEGER
route_source_value	No	VARCHAR(50)
dose_unit_source_value	No	VARCHAR(50)

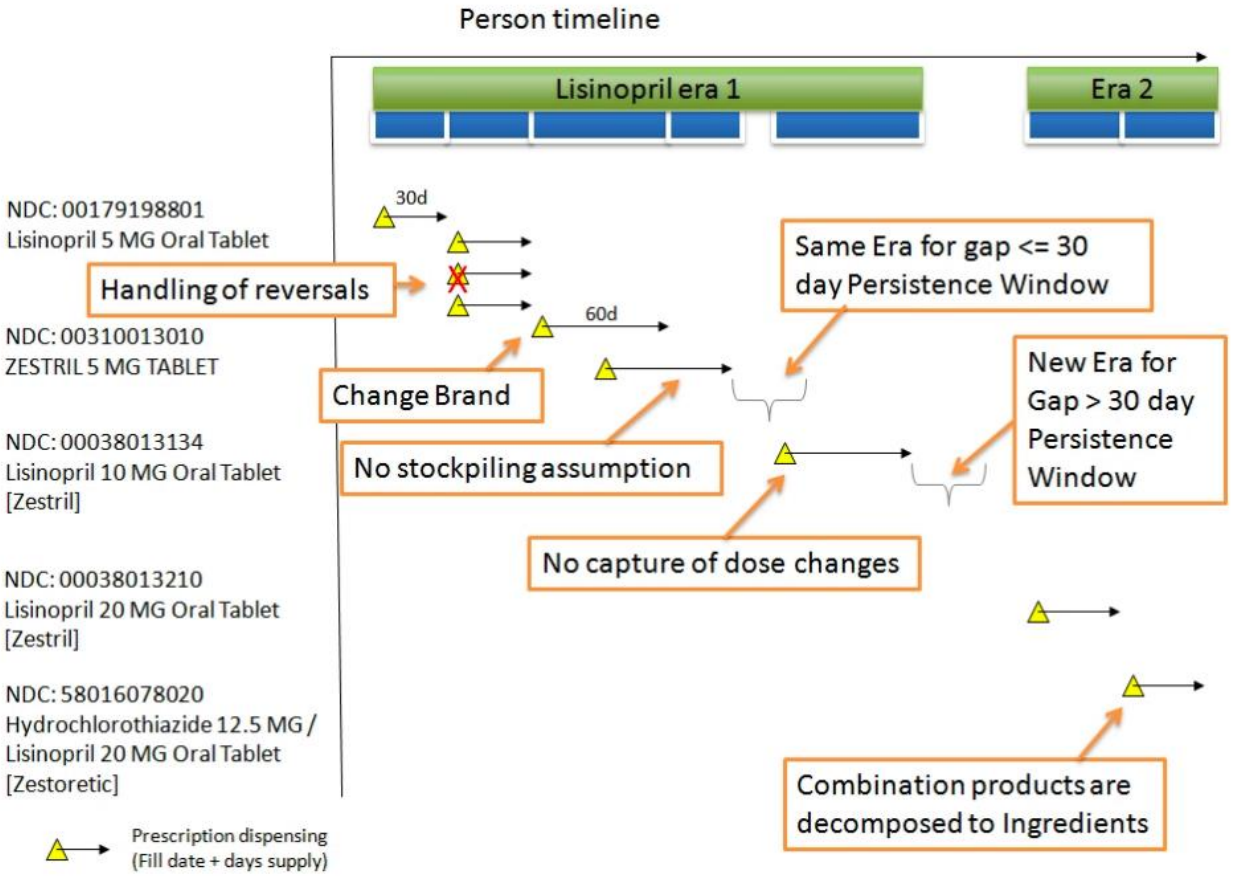
- 약물 처방 및 투약 기록
- 약품명, 처방일, 총수량, 투여경로

  (경구, 주사, 외용)
- 표준용어: RxNorm, RxNorm

  Extension

DRUG ERA

DRUG_ERA



분당서울대 유수영 교수님 Slide

DRUG ERA

DRUG_ERA

field	required	type
drug_era_id	Yes	INTEGER
person_id	Yes	INTEGER
drug_concept_id	Yes	INTEGER
drug_era_start_datetime	Yes	DATE
drug_era_end_datetime	Yes	DATE
drug_exposure_count	No	INTEGER
gap_days	No	INTEGER

- Ingredient level의 연속적인 약물 노출 기간 추론 데이터
- DRUG_EXPOSURE 테이블로부터 표준화된 로직을 적용하여 자동 계산(예, 30일 persistence window 허용 등)

PROCEDURE OCCURRENCE

PROCEDURE_OCCURRENCE

field	required	type
procedure_occurrence_id	Yes	INTEGER
person_id	Yes	INTEGER
procedure_concept_id	Yes	INTEGER
procedure_date	No	DATE
procedure_datetime	Yes	DATETIME
procedure_type_concept_id	Yes	INTEGER
modifier_concept_id	Yes	INTEGER
quantity	No	INTEGER
provider_id	No	INTEGER
visit_occurrence_id	No	INTEGER
visit_detail_id	No	INTEGER
procedure_source_value	No	VARCHAR(50)
procedure_source_concept_id	Yes	INTEGER
modifier_source_value	No	VARCHAR(50)

- 수술, 처치, 검사처방 등
- 표준용어: SNOMED CT
- 표준 용어 매핑 이슈 존재

MEASUREMENT

MEASUREMENT

field	required	type
measurement_id	Yes	INTEGER
person_id	Yes	INTEGER
measurement_concept_id	Yes	INTEGER
measurement_date	No	DATE
measurement_datetime	Yes	DATETIME
measurement_time	No	VARCHAR(10)
measurement_type_concept_id	Yes	INTEGER
operator_concept_id	No	INTEGER
value_as_number	No	FLOAT
value_as_concept_id	No	INTEGER
unit_concept_id	No	INTEGER
range_low	No	FLOAT
range_high	No	FLOAT
provider_id	No	INTEGER
visit_occurrence_id	No	INTEGER
visit_detail_id	No	INTEGER
measurement_source_value	No	VARCHAR(50)
measurement_source_concept_id	Yes	INTEGER
unit_source_value	No	VARCHAR(50)
value_source_value	No	VARCHAR(50)

- 수치형/범주형의 구조화된 검사결과, 임상관찰기록, 병리보고서 결과 등
- 표준용어: LOINC, SNOMED CT
- 표준 용어 매핑 이슈 존재

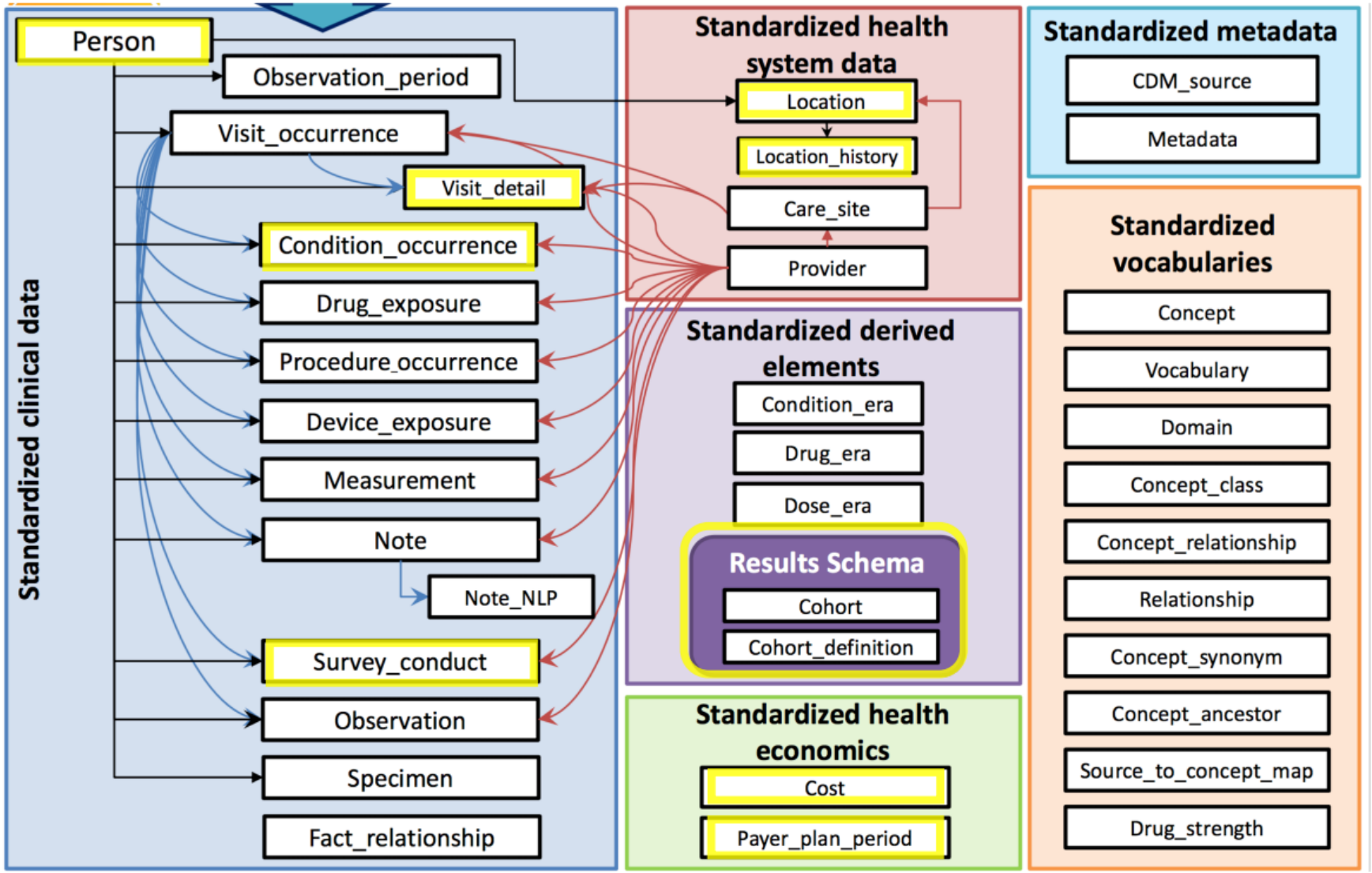
OBSERVATION

OBSERVATION

field	required	type
observation_id	Yes	INTEGER
person_id	Yes	INTEGER
observation_concept_id	Yes	INTEGER
observation_date	No	DATE
observation_datetime	Yes	DATETIME
observation_type_concept_id	Yes	INTEGER
value_as_number	No	FLOAT
value_as_string	No	VARCHAR(60)
value_as_concept_id	No	INTEGER
qualifier_concept_id	No	INTEGER
unit_concept_id	No	INTEGER
provider_id	No	INTEGER
visit_occurrence_id	No	INTEGER
visit_detail_id	No	INTEGER
observation_source_value	No	VARCHAR(50)
observation_source_concept_id	Yes	INTEGER
unit_source_value	No	VARCHAR(50)
qualifier_source_value	No	VARCHAR(50)
observation_event_id	No	INTEGER
obs_event_field_concept_id	Yes	INTEGER
value_as_datetime	No	INTEGER

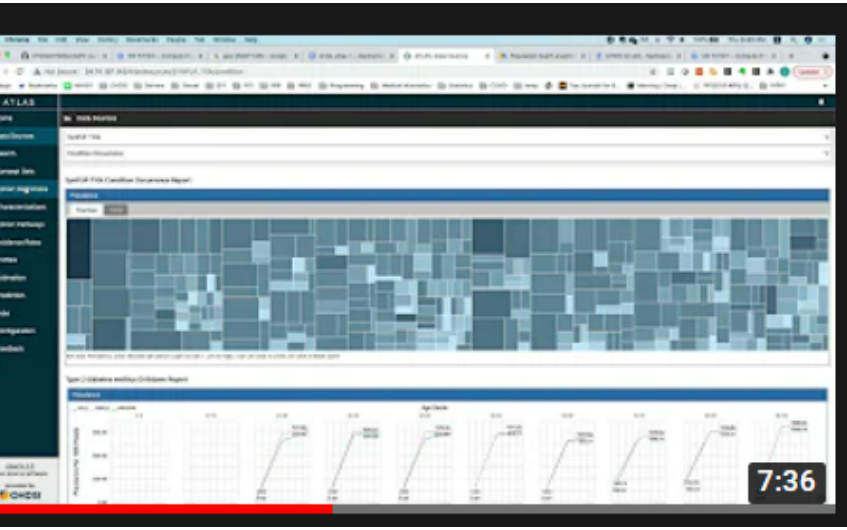
- 가족력/과거력, 흡연, 음주, 문진 결과, 평가지 등 데이터
- 기타 다른 테이블에 저장할 수 없는 임상 데이터
- 표준용어: SNOMED CT, LOINC

OMOP Version 6 Key Domains



분당서울대 유수영 교수님 Slide

Do Phenotyping!



구글클라우드(GCP) 이용 OHDSI ATLAS 인스턴스 실행

조회수 82회 • 9개월 전

S Seng Chan You

구글클라우드(GCP)를 이용하여 OHDSI ATLAS를 실행하고, SynPUF 합성 데이터 기반으로 다양한 실습/실험을 할 수 있는 방법 안내 ...

https://youtu.be/yXLd6DCp26A