CDM인프라 및 연구방법 개론

연세대학교 의생명시스템정보학교실 유승찬

chandryou@yuhs.ac

Contents

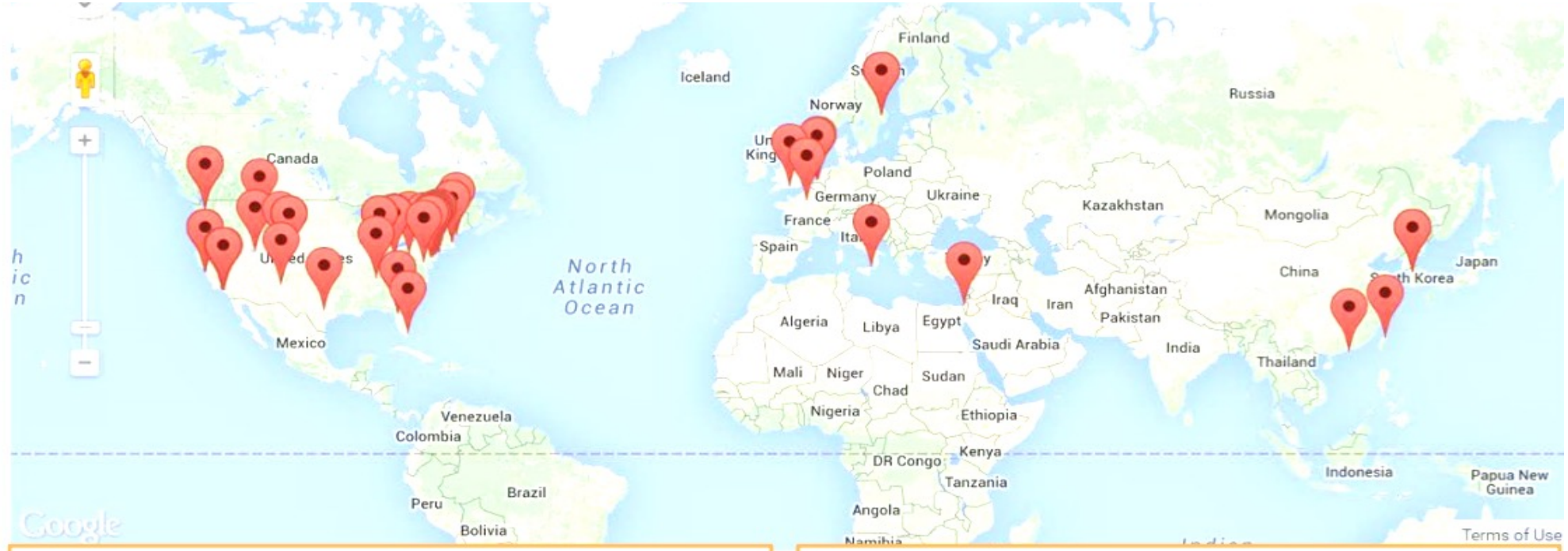
- Common Data Model (CDM) / OHDSI
- Open-source and Characterization
- Population-level estimation: Large-scale Evidence Generation and Evaluation in a Network of Databases
- Patient-Level Prediction

Contents

- Common Data Model (CDM) / OHDSI
- Characterization
- Population-level estimation: Large-scale Evidence Generation and Evaluation in a Network of Databases
- Patient-Level Prediction

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:

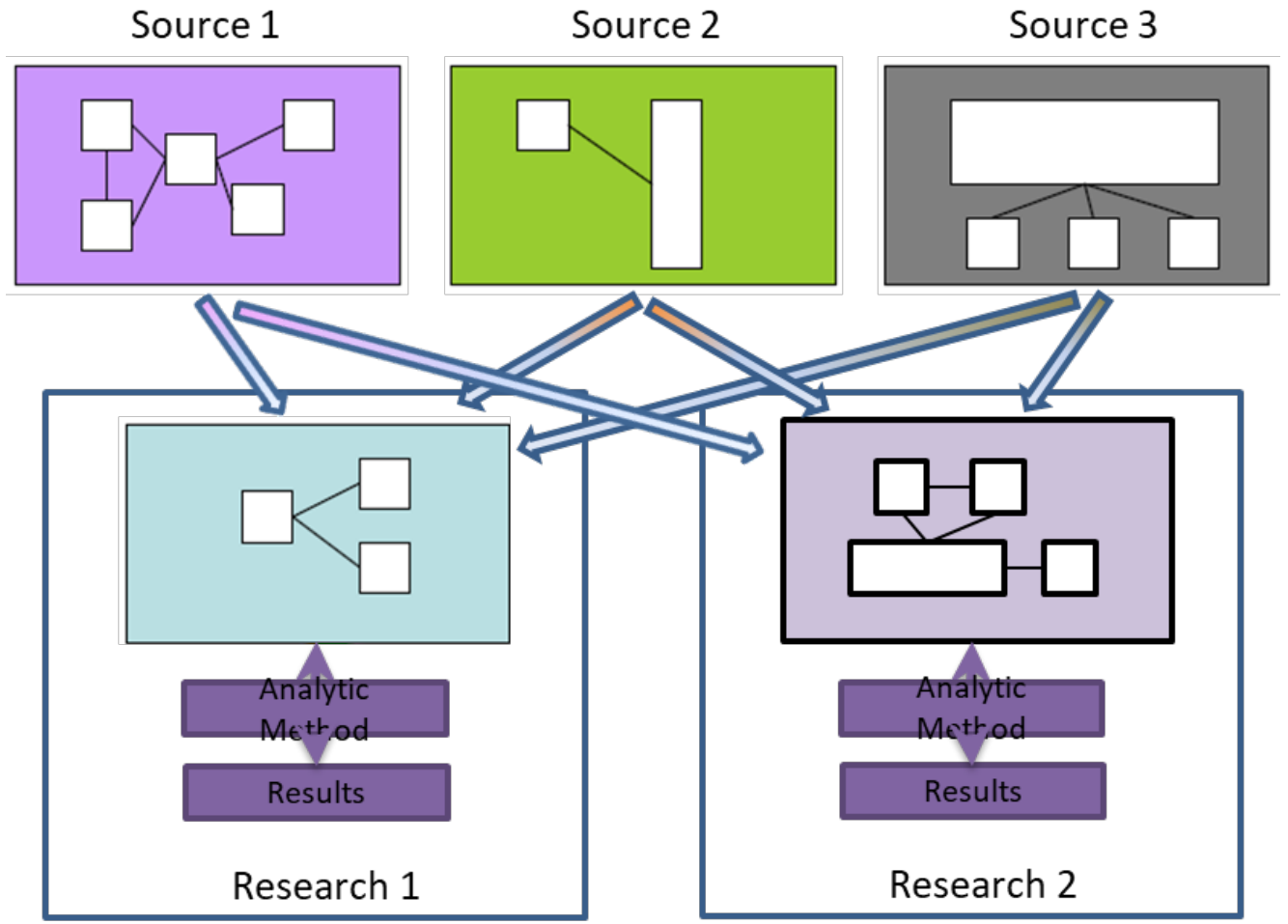
- >40 databases standardized to OMOP common data model
- >500 million patients

https://www.ohdsi.org/

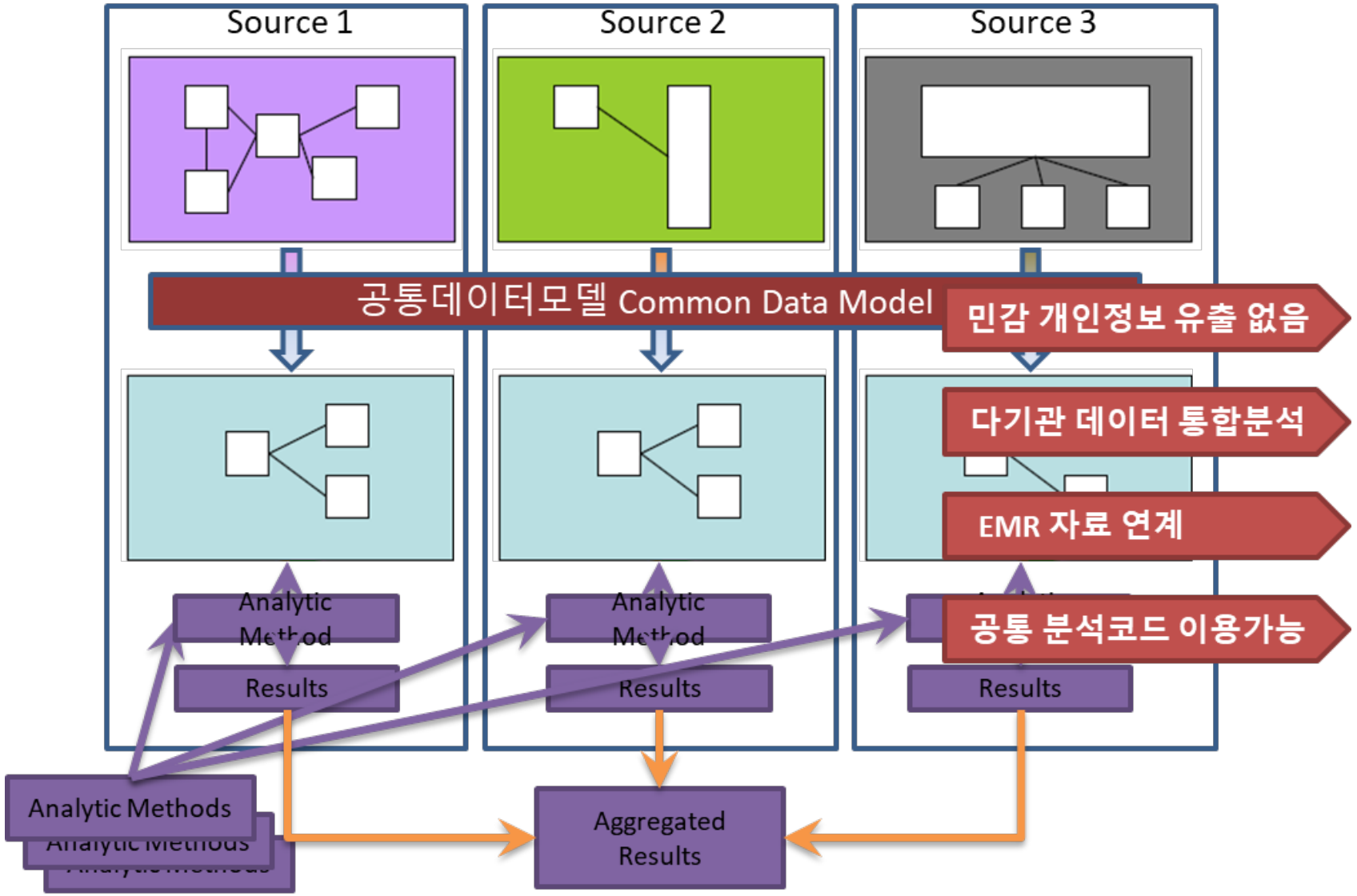
Why Common Data Model (CDM)?

기존의 다기관 연구방법

연구 수행 때마다 데이터 모델을 맞추는 변환 작업을 수행해야 함



CDM in Distributed Research Network



분산연구망 및 공통데이터모델 CDM (Common Data Moldel)

기술적문제

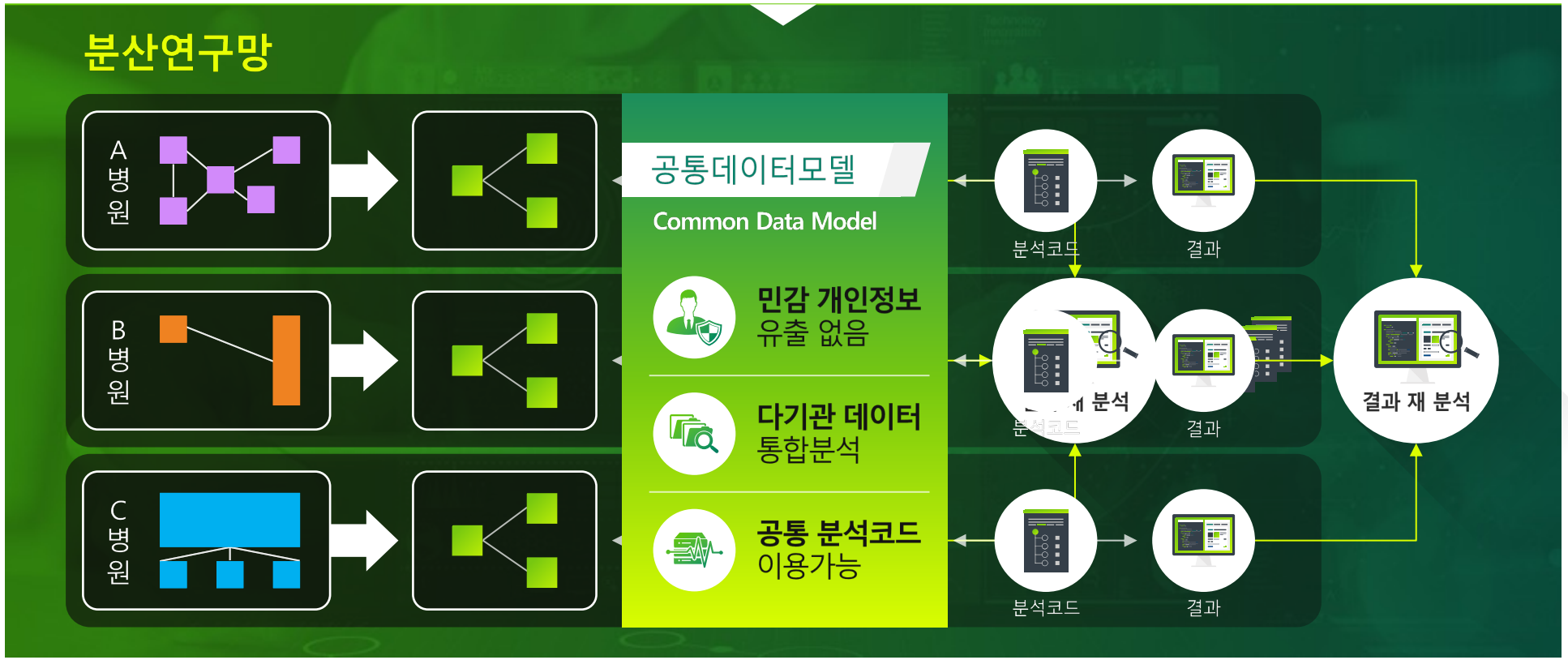
- 데이터 구조 / 형식의 **이질성**
- 데이터의 **질과 양**
- 기술적 **한계**

법적문제

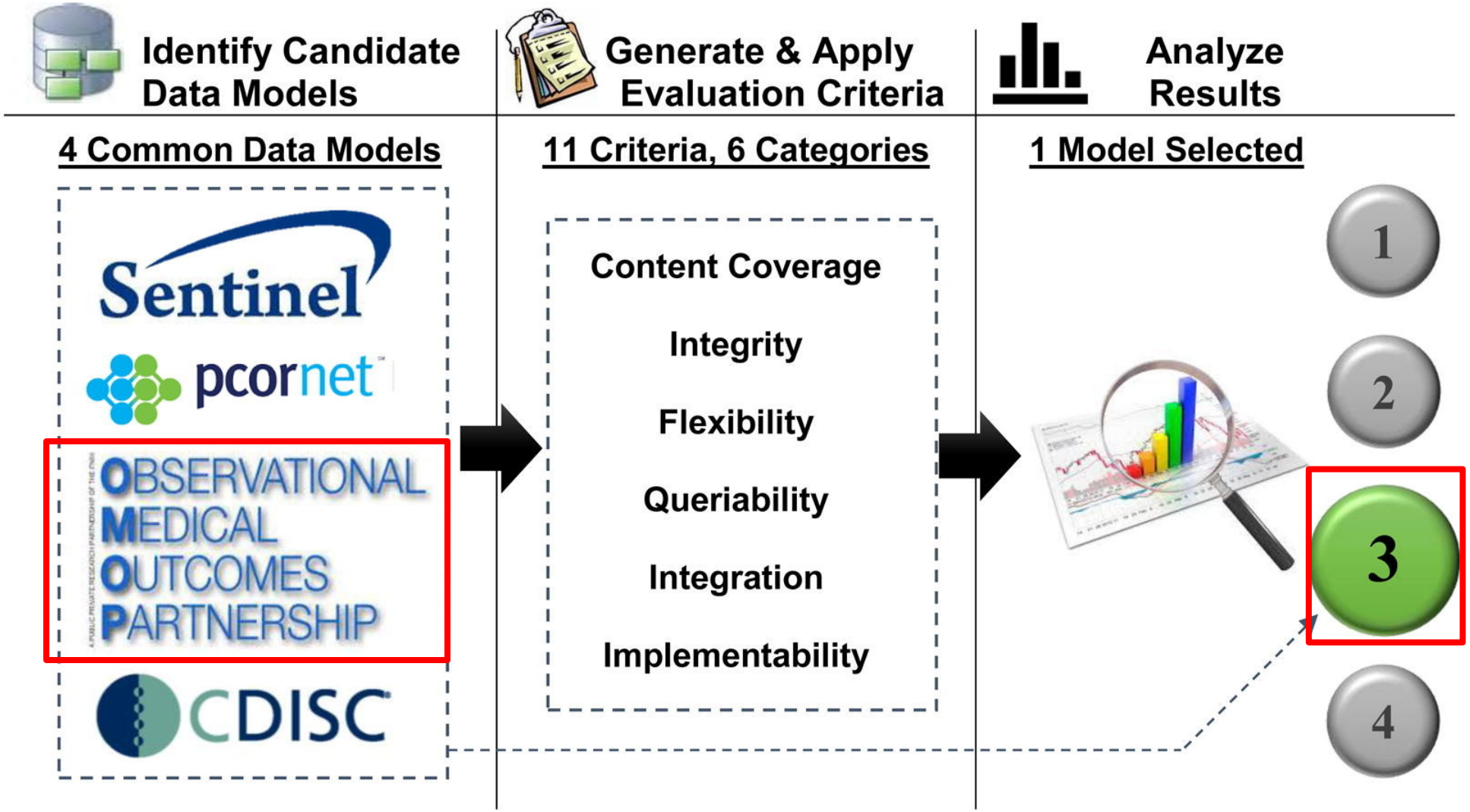
- **개인정보보호법**
- 기관의 **허락**

인간본성/감정적문제

- 자신에게 **불리하게 사용될지 모른다는 두려움**
- 데이터 소유자는 자신의 자료를 **공유하고 싶어하지 않음**



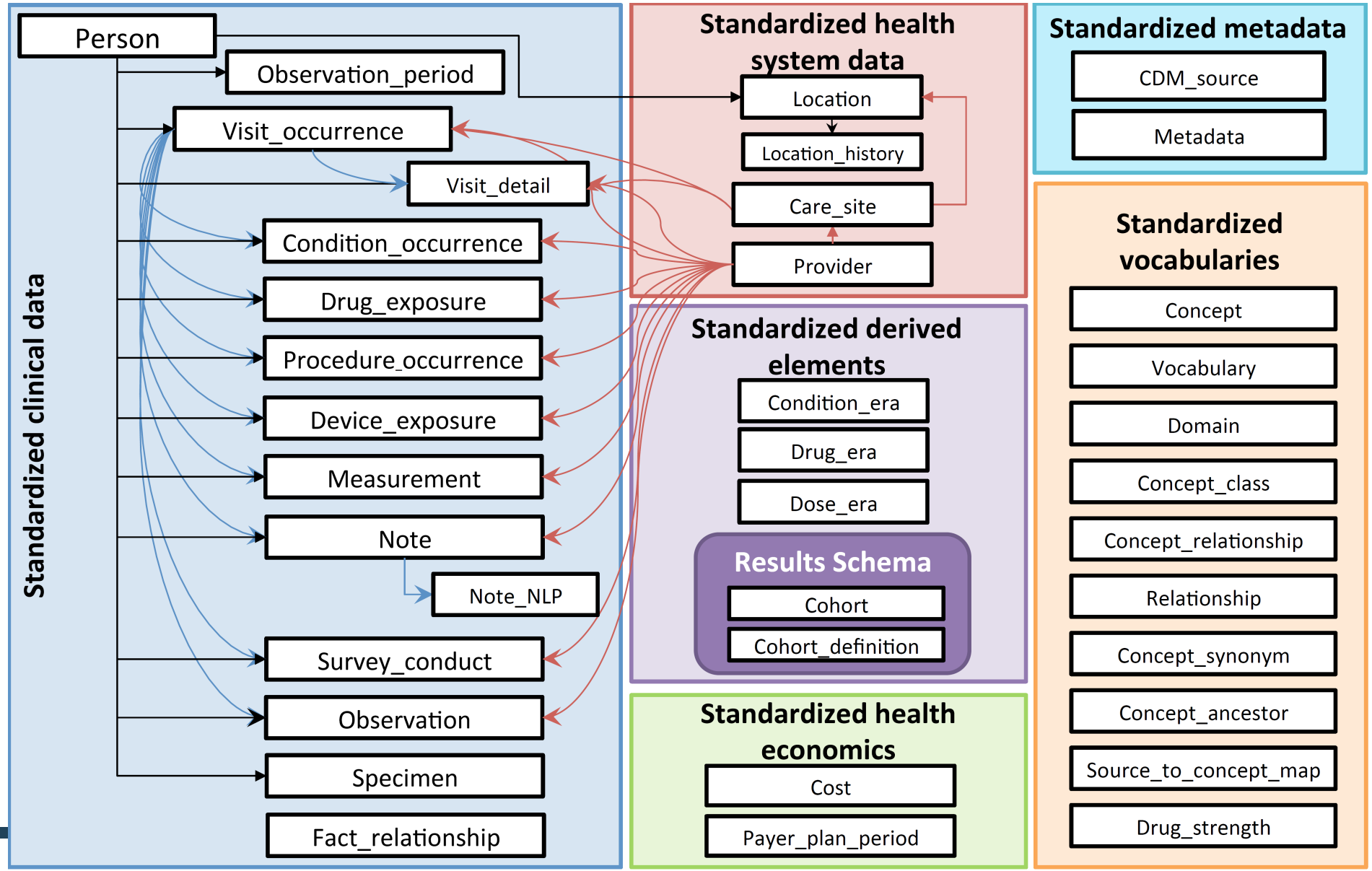
Various Common Data Models



- 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

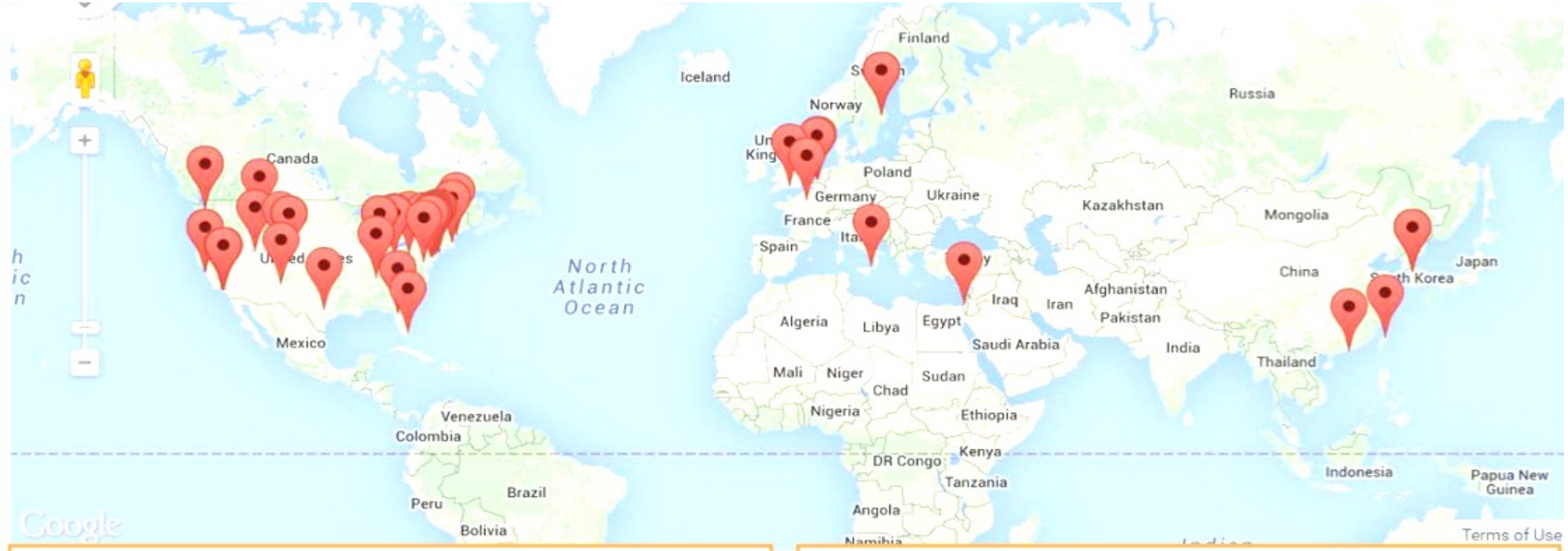
Garza, et al., JBI 2016

OMOP Common Data Model V6



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:

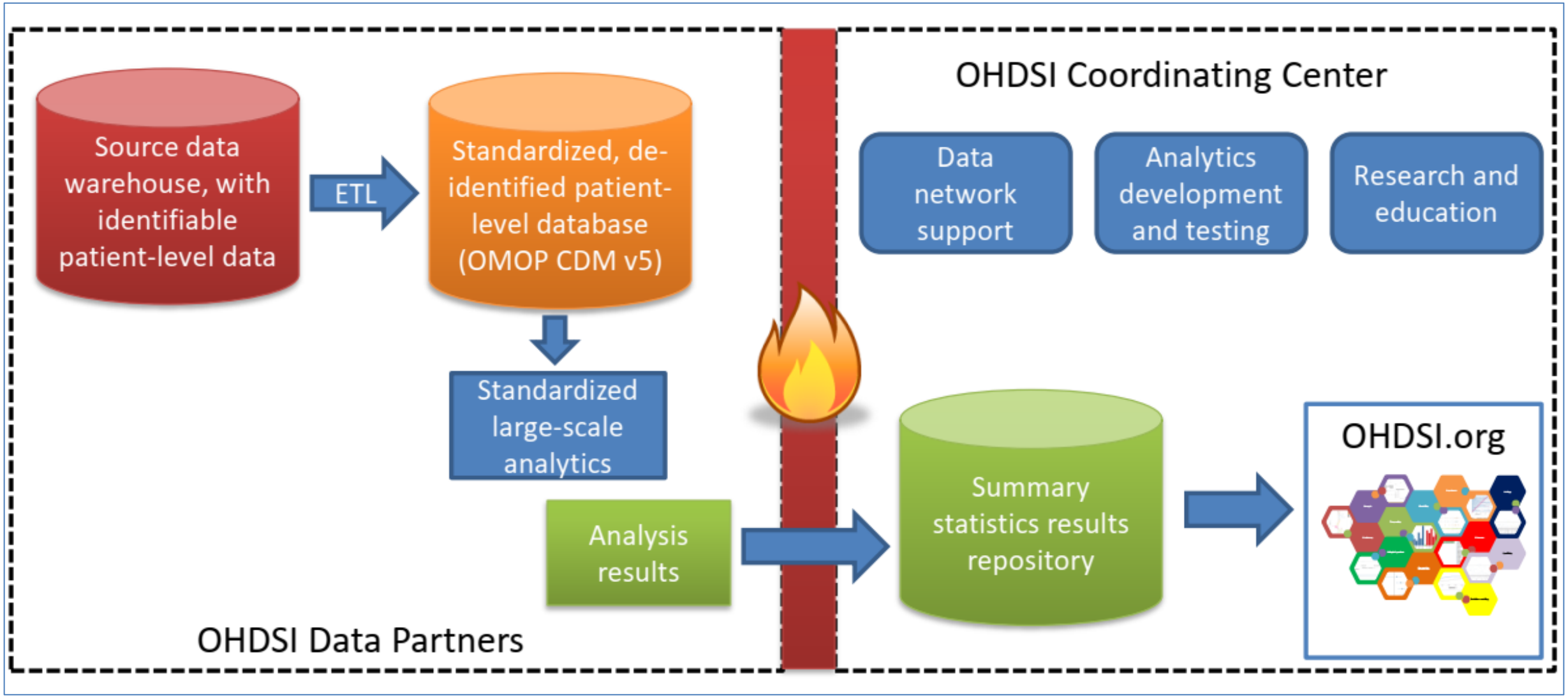
- >100 researchers in academia, industry and government
- >10 countries

OHDSI Data Network:

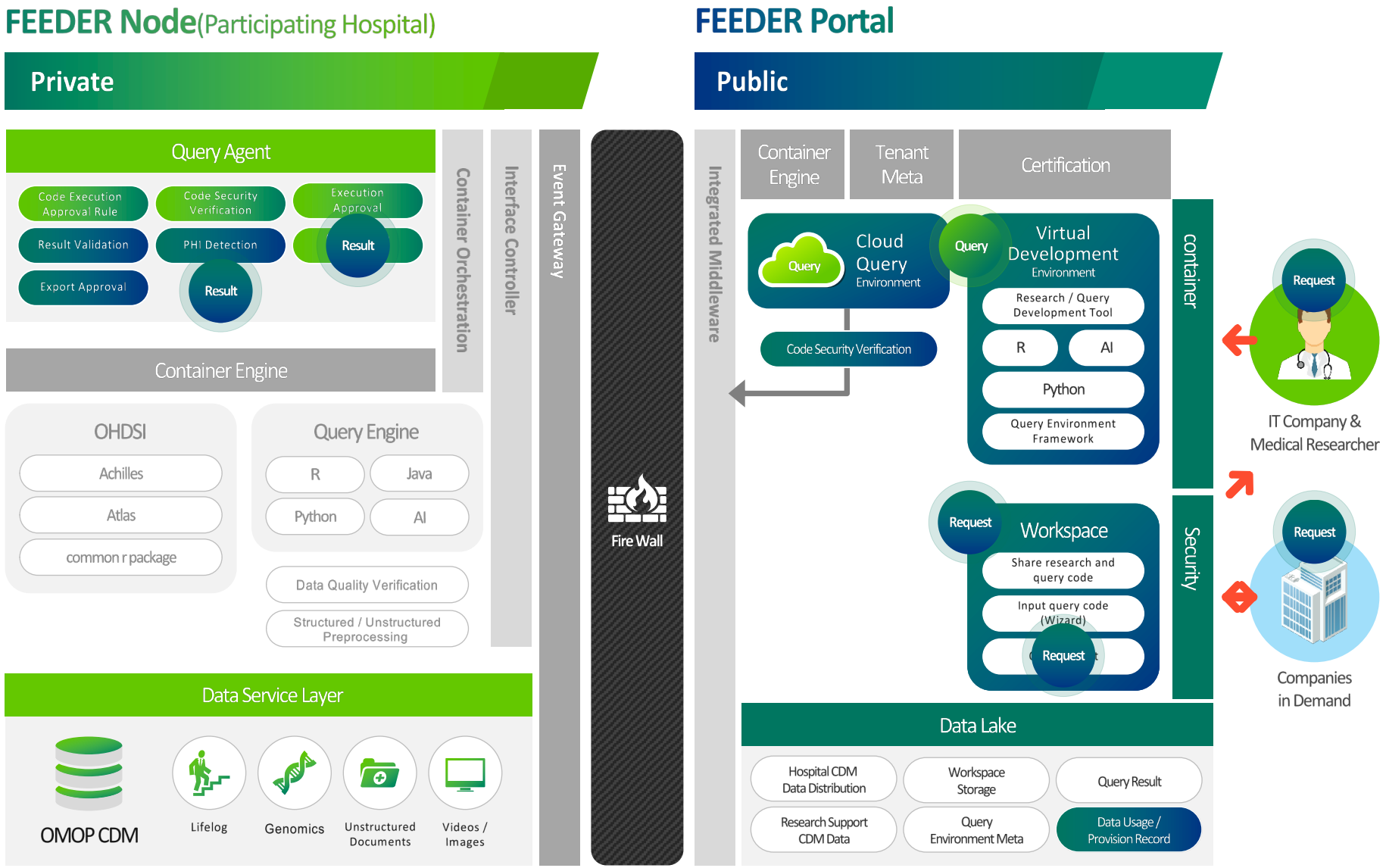
- >40 databases standardized to OMOP common data model
- >500 million patients

https://www.ohdsi.org/

How OHDSI works



How Distributed Research Platform 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

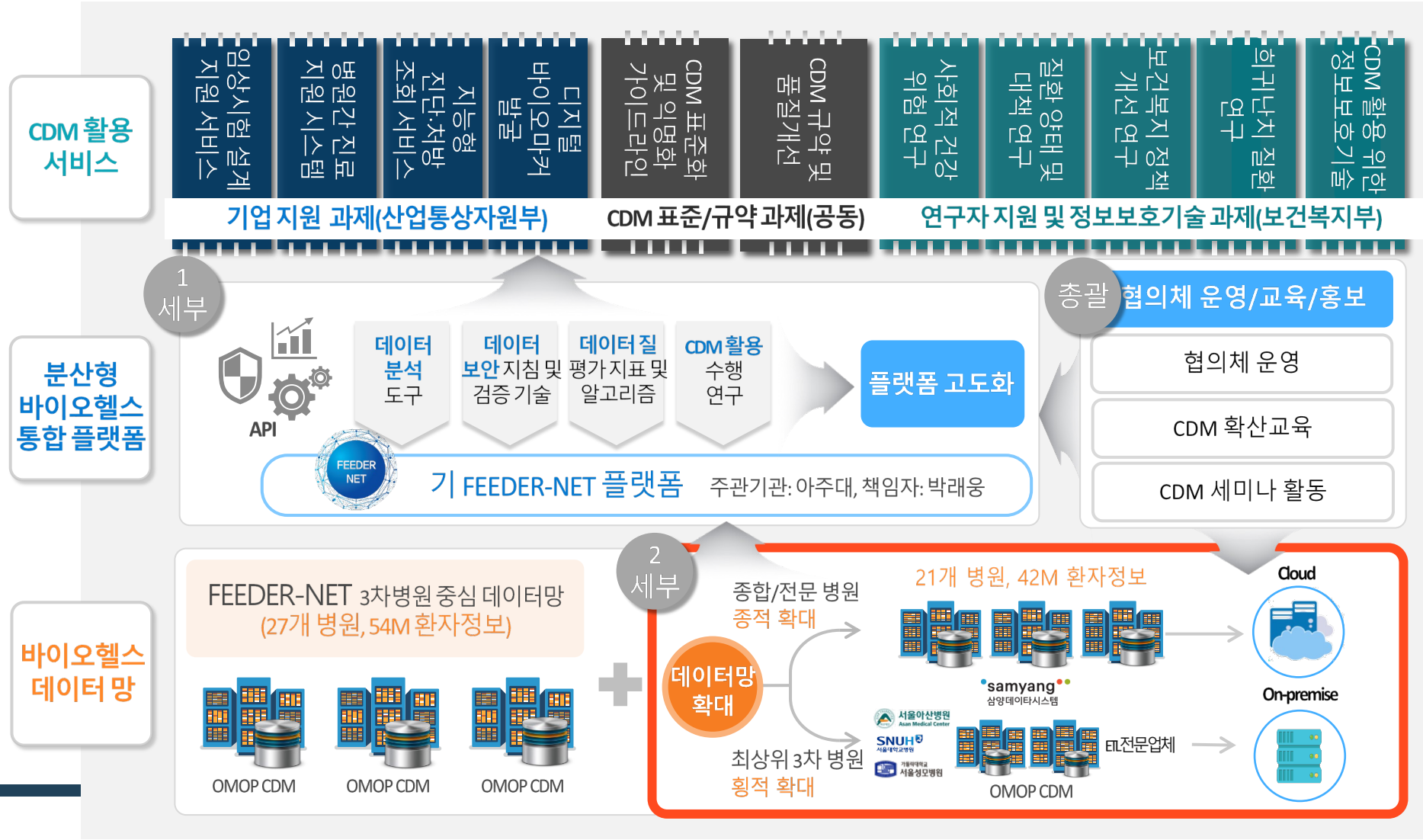
현 누적 변환 기관 수 : 32병원 (3차: 20개 / 2차: 12개)

현 누적 변환 환자 수 : 44,596,864명

공통 데이터모델 (CDM) 기반 헬스케어 융합 빅데이터 생태계 구축

산업통상자원부 CDM 사업을 통한

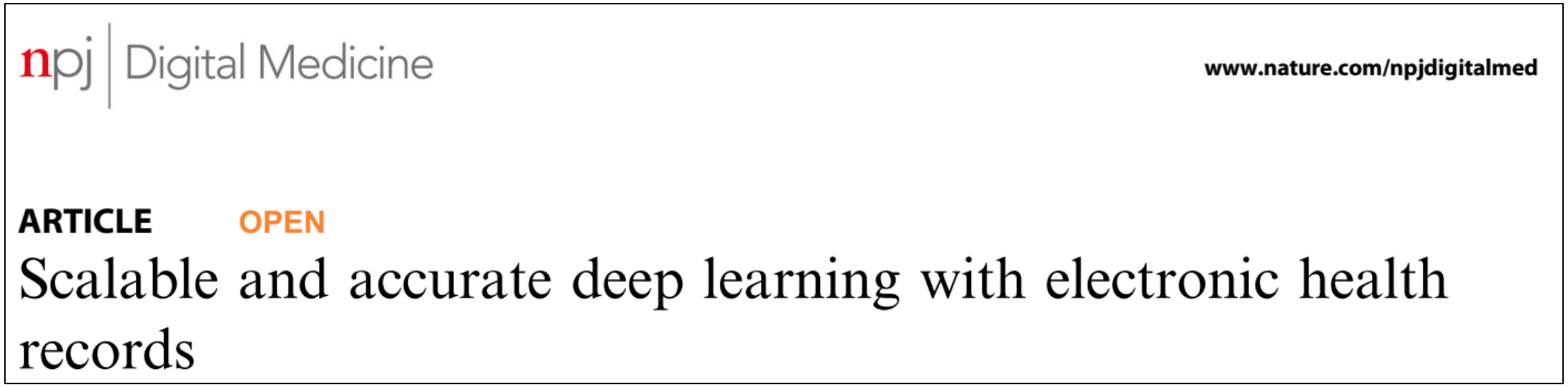
FEEDER-NET 확장



Contents

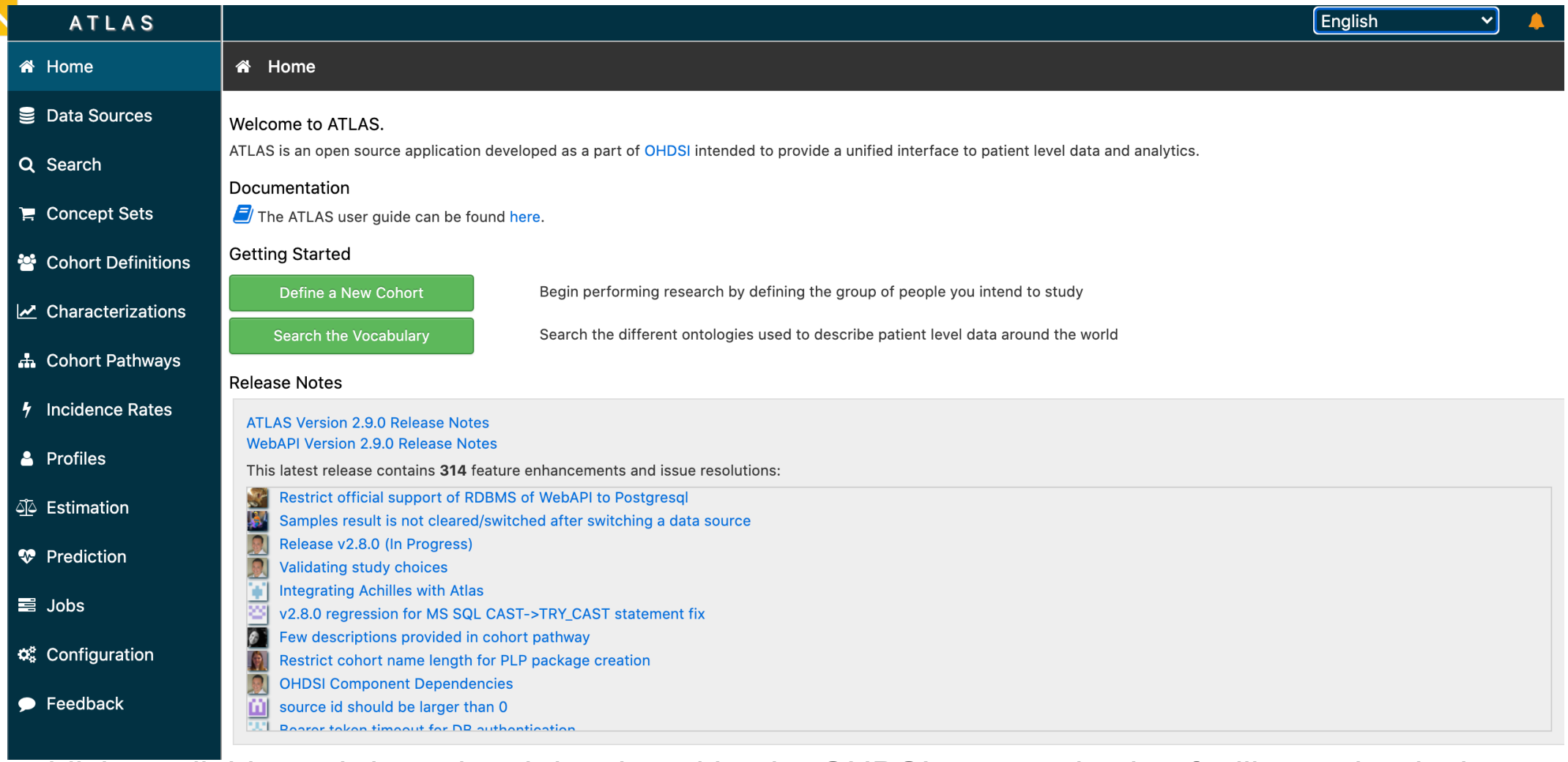
- Common Data Model (CDM) / OHDSI
- Characterization
- Population-level estimation: Large-scale Evidence Generation and Evaluation in a Network of Databases
- Patient-Level Prediction

Standardized data-based software 개발의 장점: Scalability



- It is widely held that 80% of the effort in an analytic model is preprocessing, merging, customizing, and cleaning datasets, not analyzing them for insights. **This profoundly limits the scalability of predictive models**
- It is crucial to standardize the health care data to enhance scalability of developed software

ATLAS

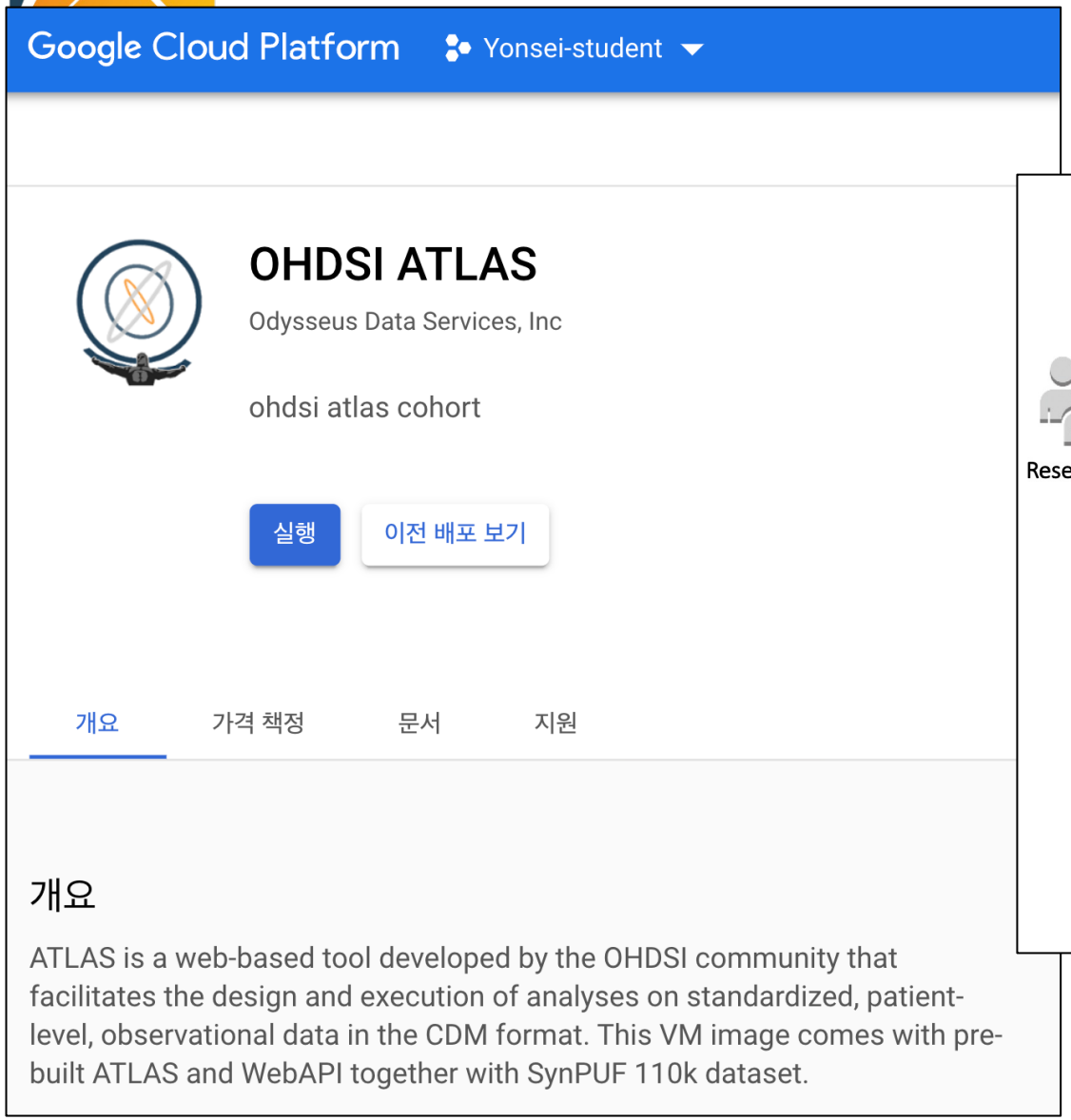


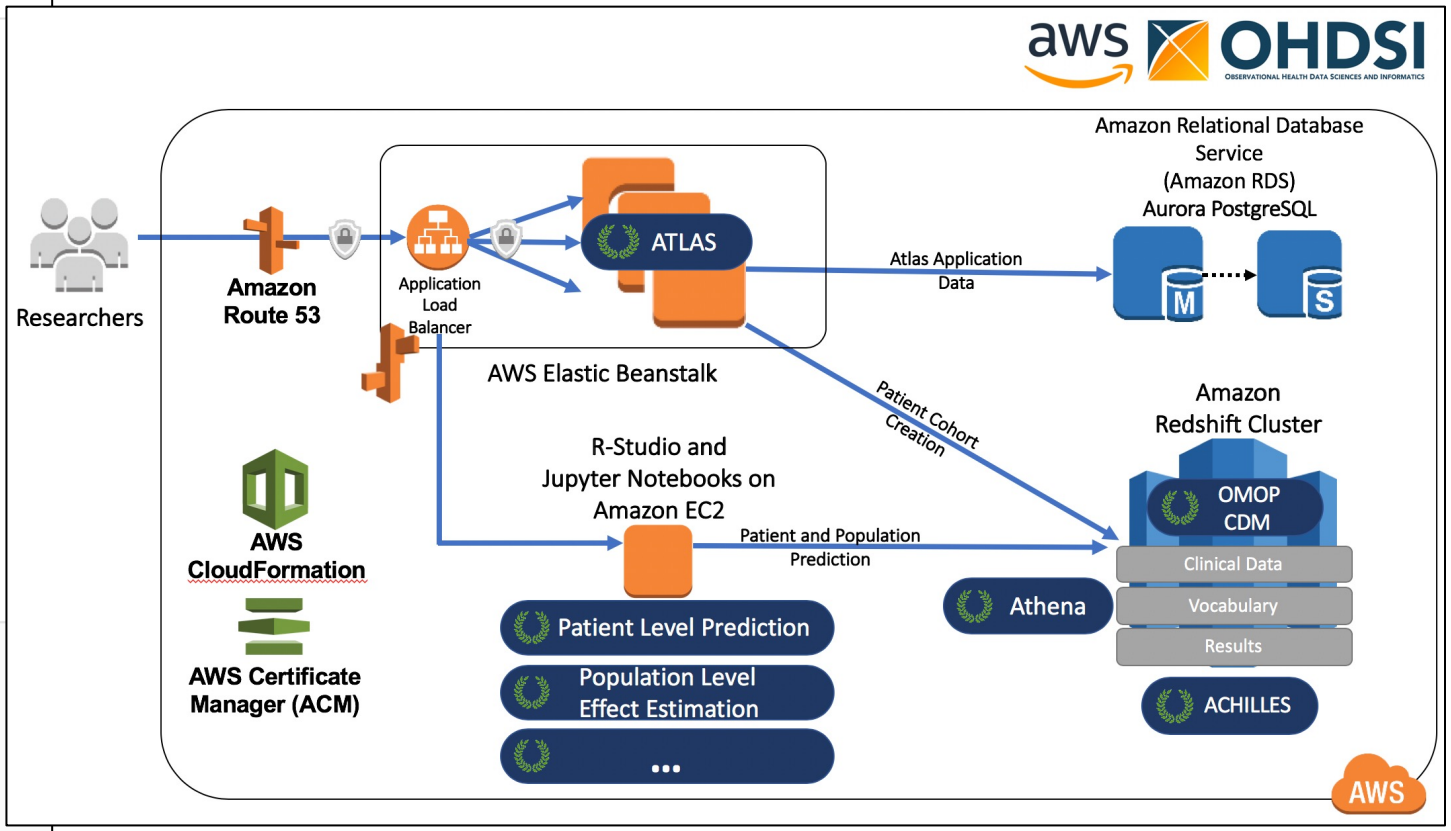
A free, publicly available, web-based tool developed by the OHDSI community that facilitates the design and execution of analyses on standardized, patient-level, observational data in the CDM format.

https://youtu.be/pMtJ3aBQ6sk

https://github.com/OHDSI/Atlas/wiki

OHDSI ATLAS on GCP or AWS

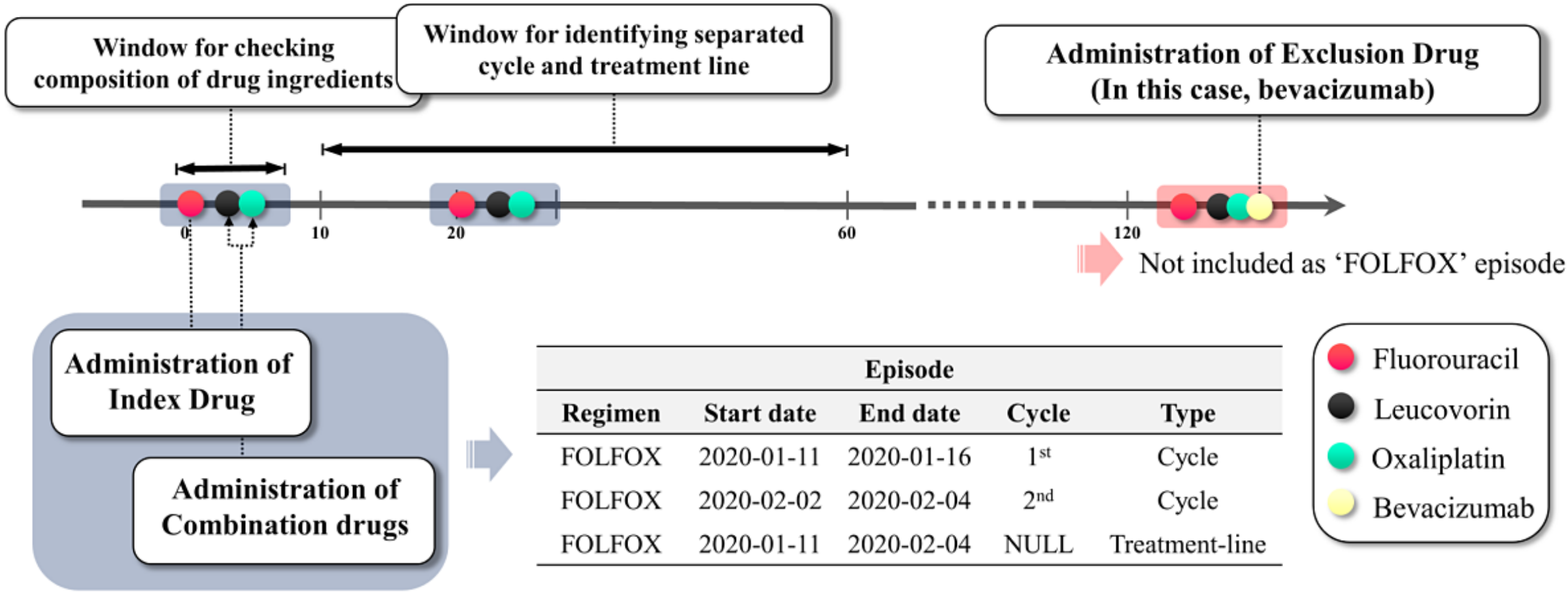




https://youtu.be/yXLd6DCp26A

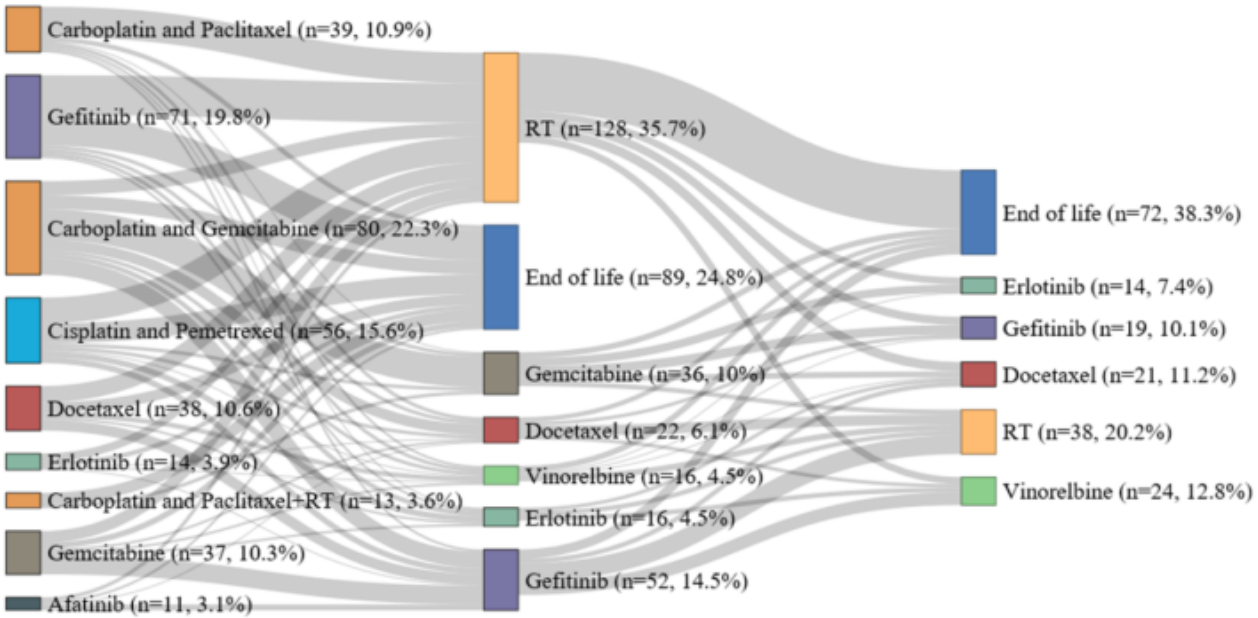
https://github.com/OHDSI/OHDSIonAWS

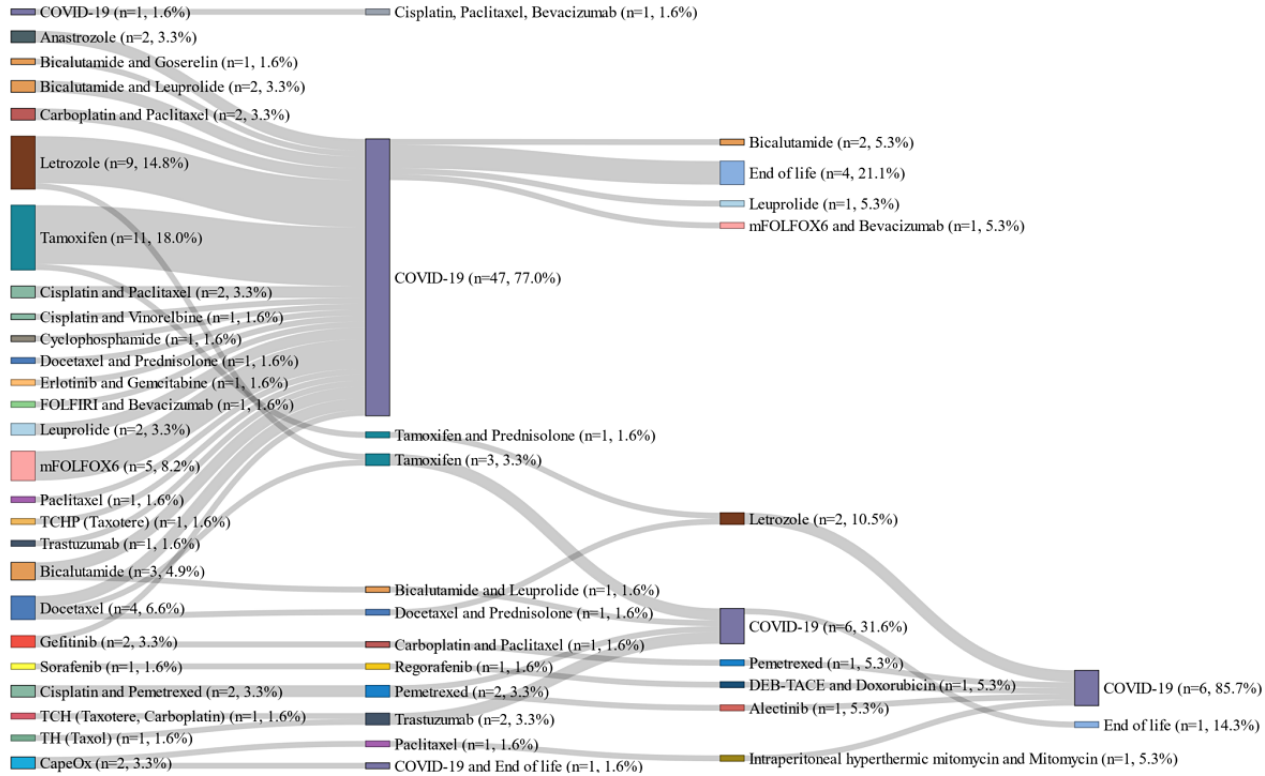
Characterization of anticancer treatment trajectory



Episode				
Regimen	**Start date**	**End date**	**Cycle**	**Type**
FOLFOX	2020-01-11	2020-01-16	1st	Cycle
FOLFOX	2020-02-02	2020-02-04	2nd	Cycle
FOLFOX	2020-01-11	2020-02-04	NULL	Treatment-line

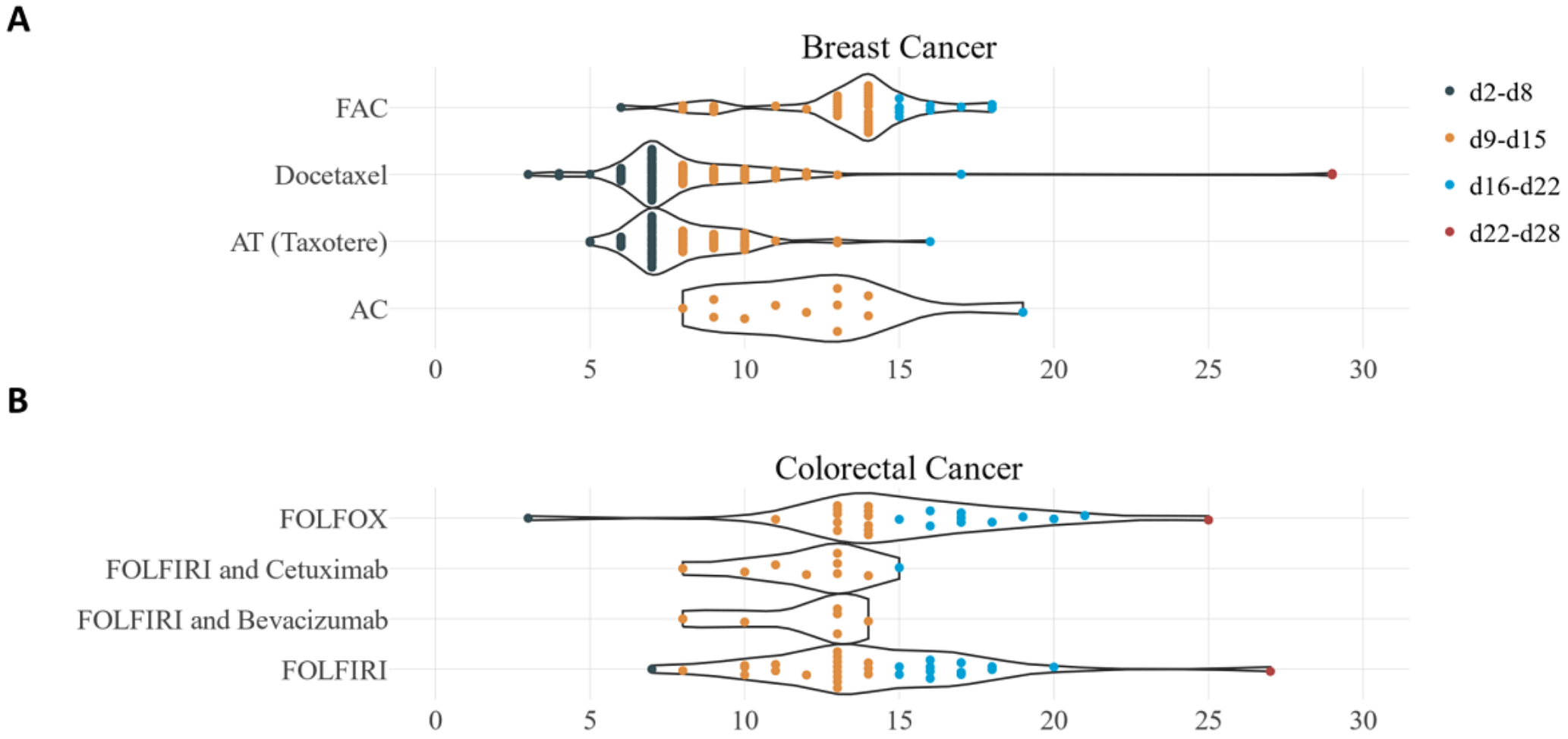
Characterization of anticancer treatment trajectory



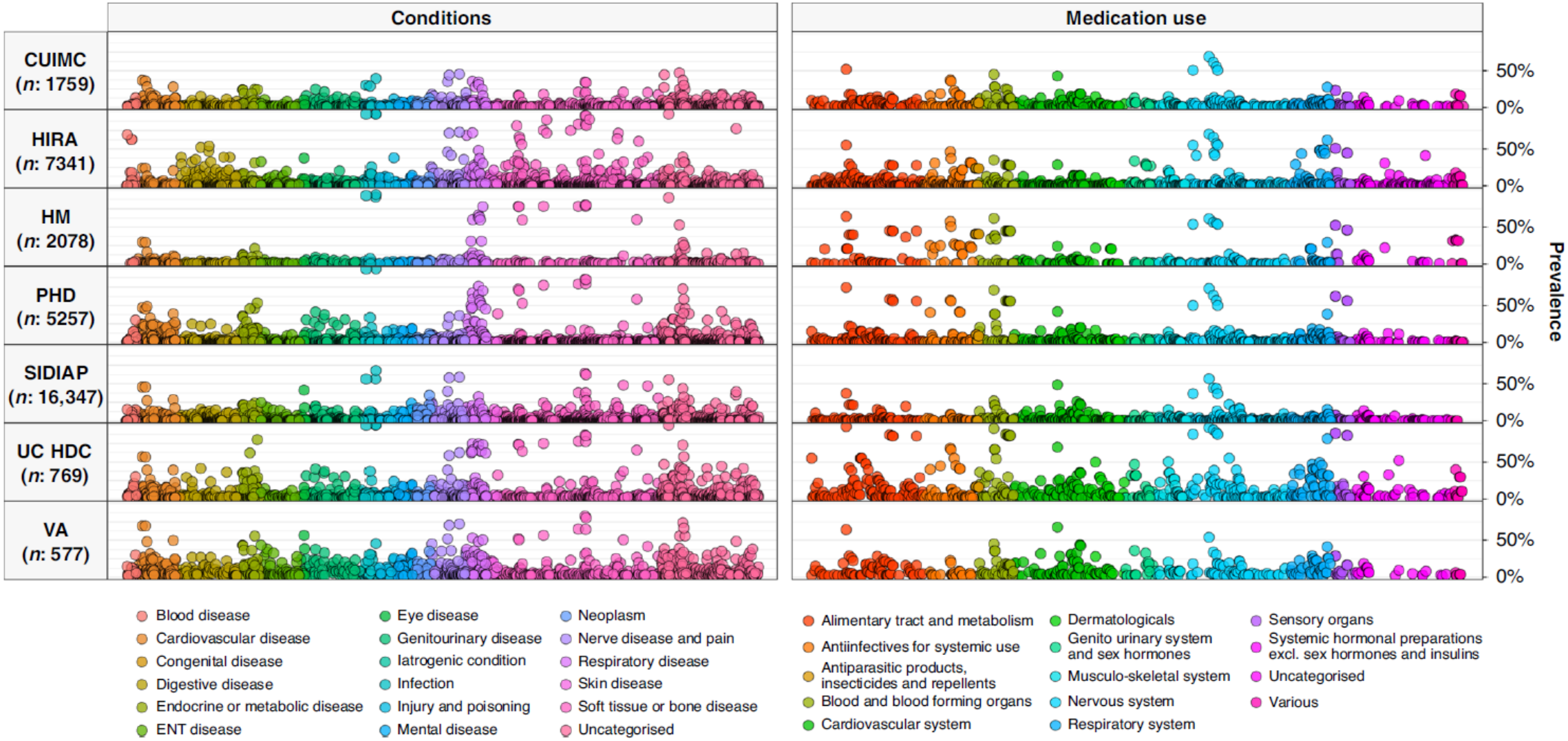


HJeon, SCYou, et al., JMIR MI, 2021

Neutropenia onset timing across various treatment

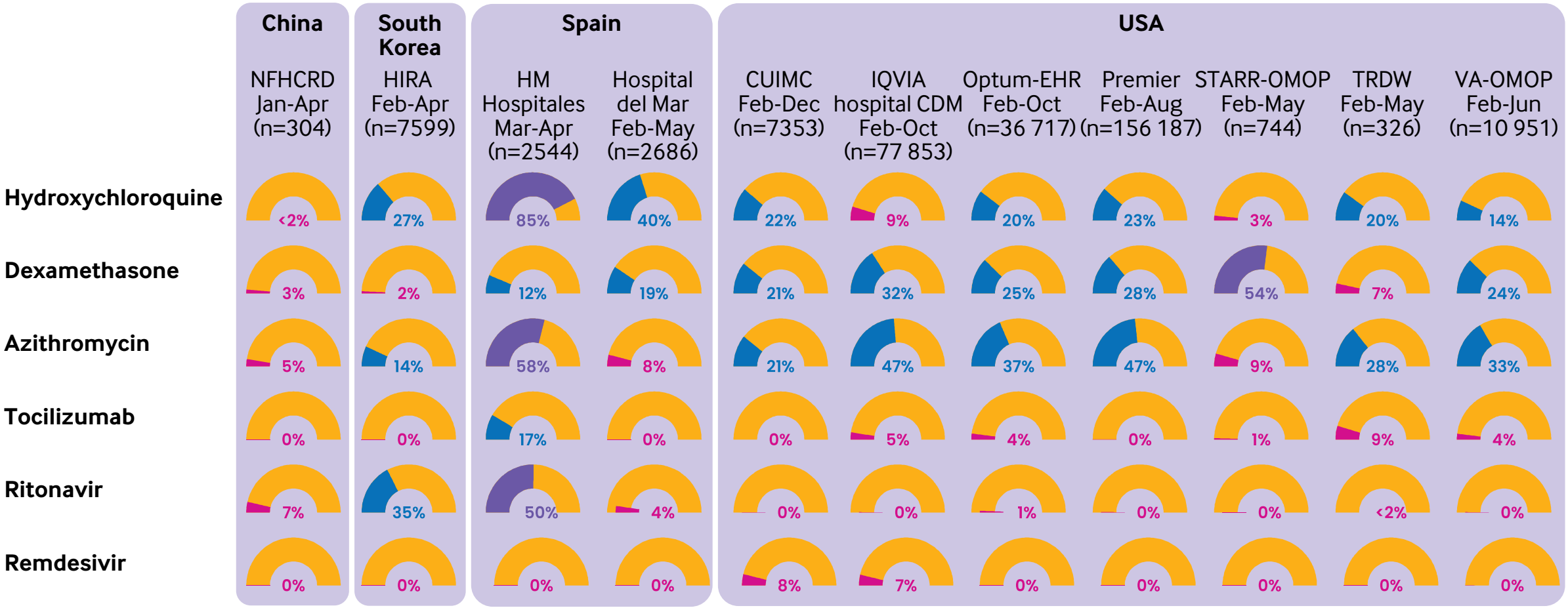


Deep phenotyping of 34,128 patients hospitalized with COVID-19 in an international network study



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

Use of repurposed and adjuvant drugs in hospital patients with covid-19: multinational network cohort study

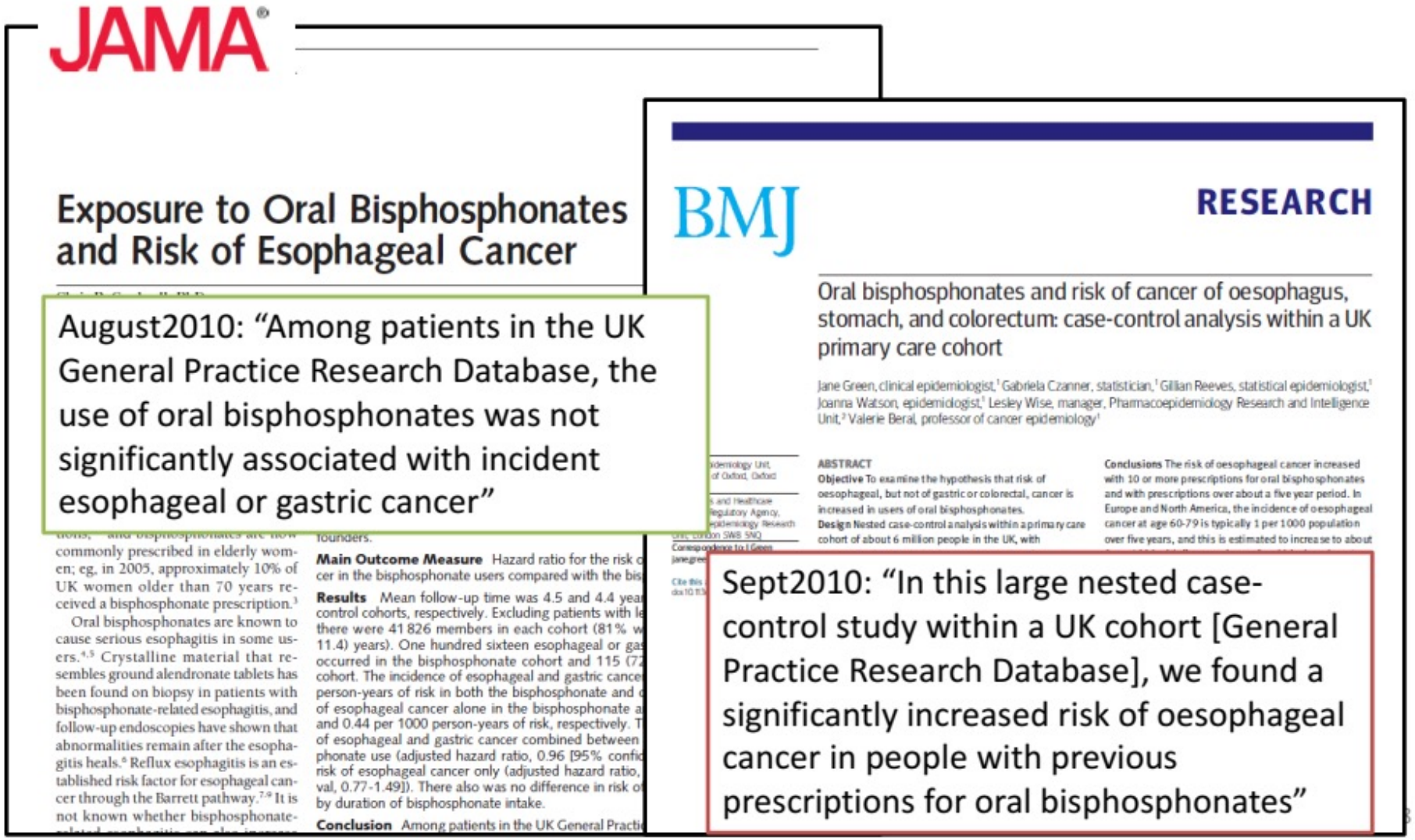


Prats-Uribe, et al., BMJ, 2021

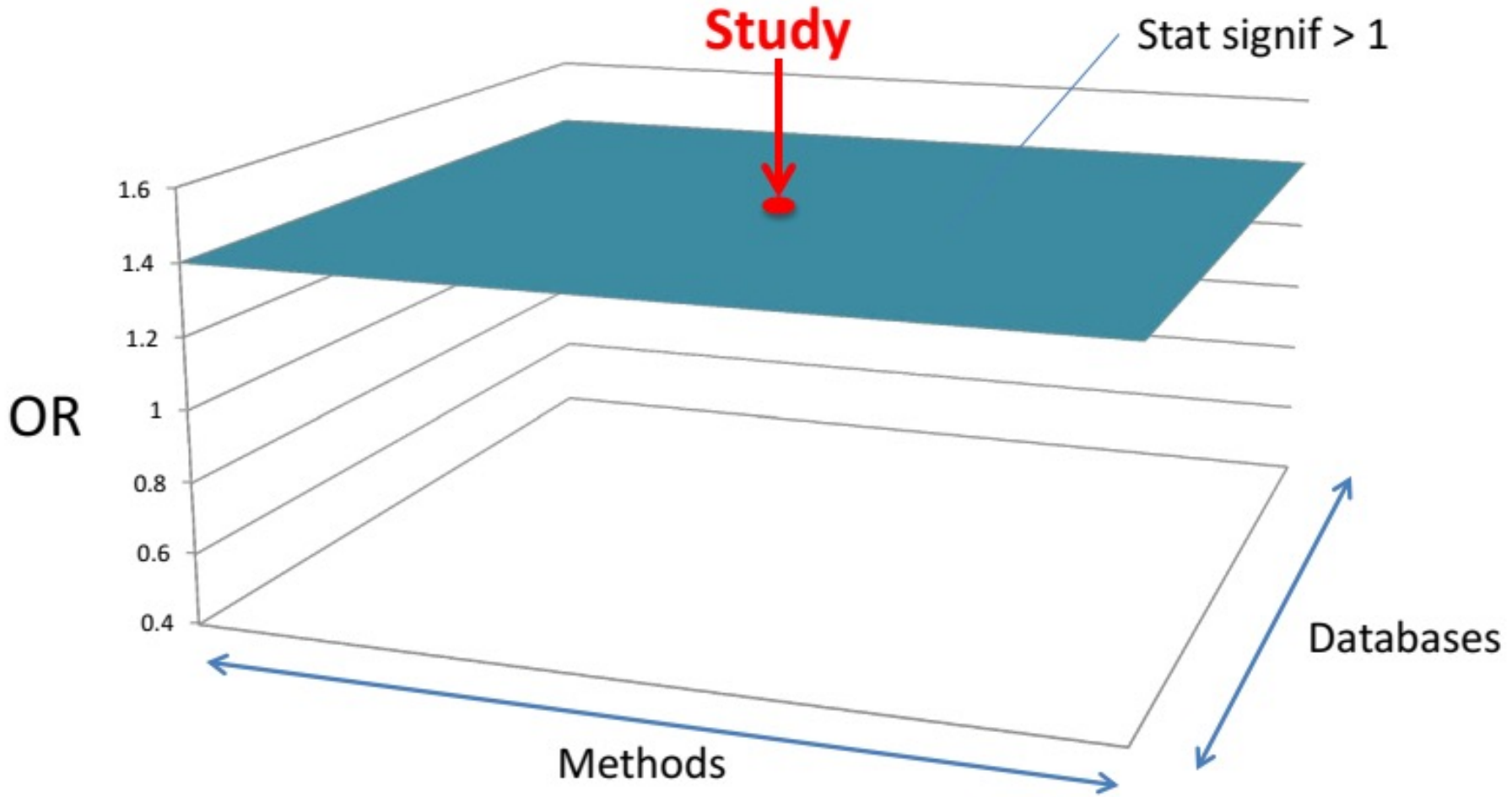
Contents

- Common Data Model (CDM) / OHDSI
- Characterization
- Population-level estimation: Large-scale Evidence Generation and Evaluation in a Network of Databases
- Patient-Level Prediction

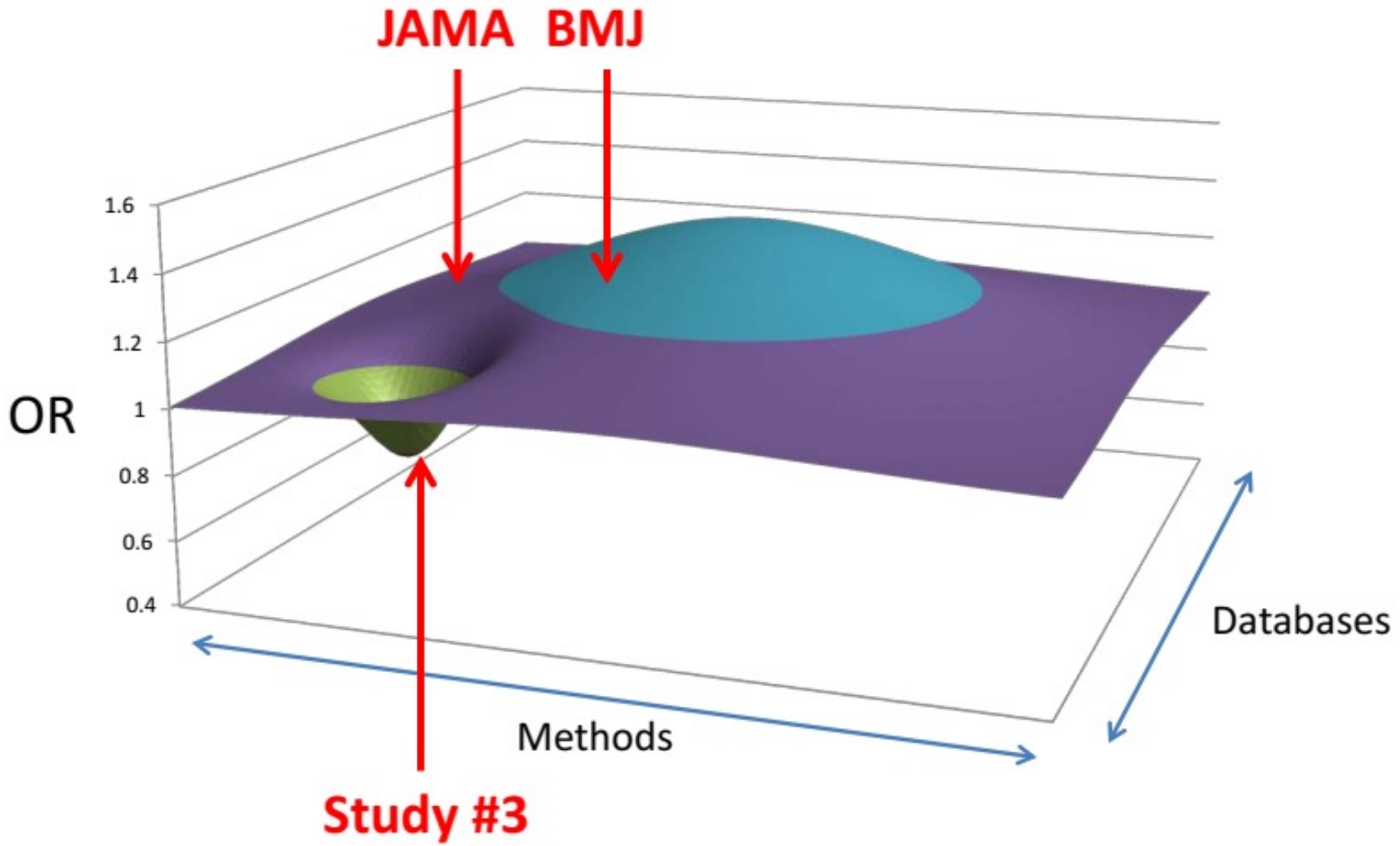
Low statistical power: current evidence from observational analyses



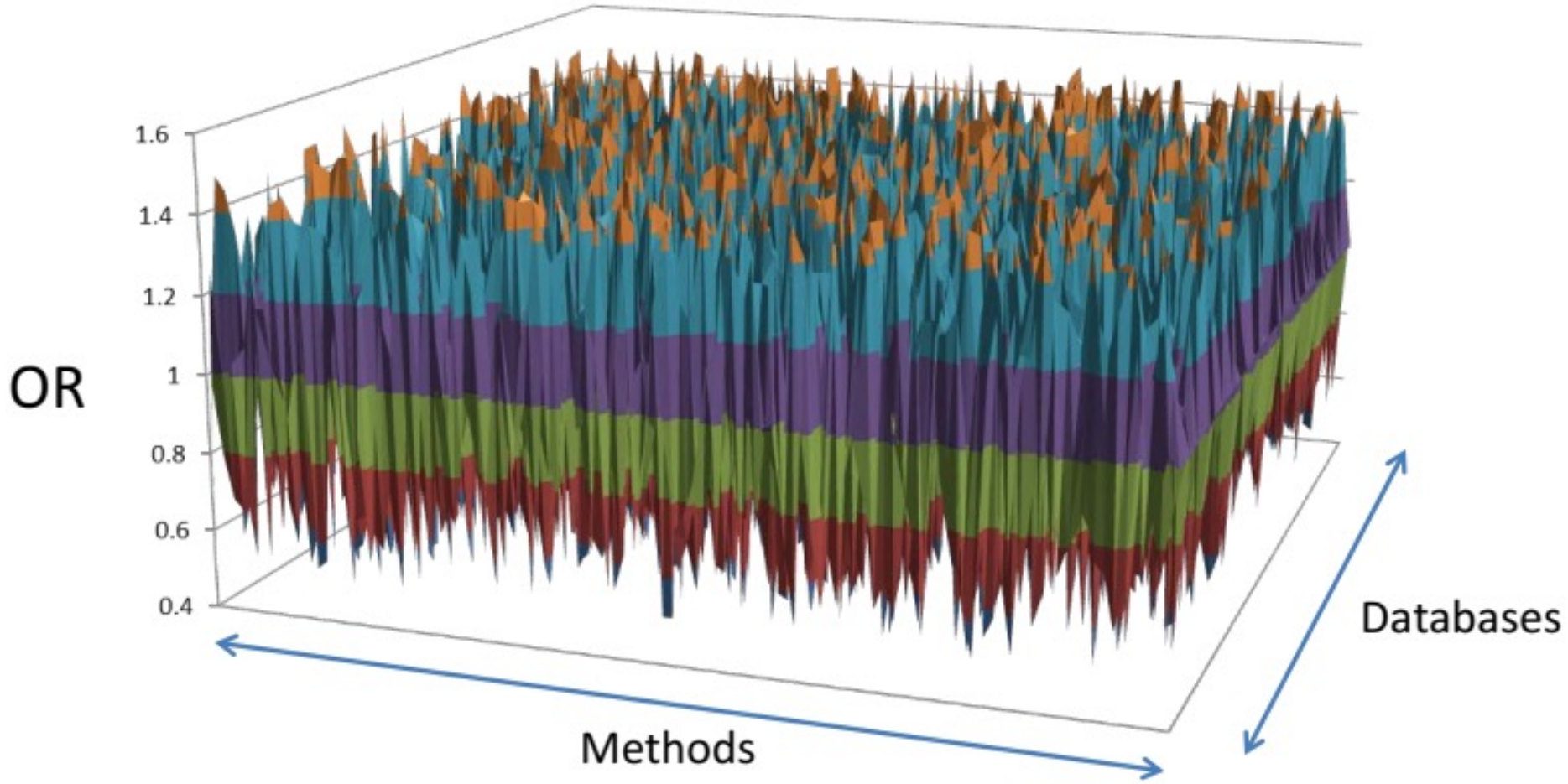
Why OHDSI: Distribution of possible results from one hypothesis



Why OHDSI: Distribution of possible results from one hypothesis



Why OHDSI: Distribution of possible results from one hypothesis



Large-scale evidence based on CDM and distributed research network

- Large-scale in terms of diverse databases
  - Heterogeneous healthcare system, enrolled patients, ethnicity, captured data
    - Korean national insurance *vs* US Medicare *vs* US private insurance data *vs* European administrative data *vs* EMR
- Large-scale in terms of analytic settings
  - Number of covariates adjusted
  - Diverse analytic settings (PS stratification *vs* PS matching)
- Large-scale in terms of number of comparisons
- Prespecified analytic process to avoid *p*-hacking

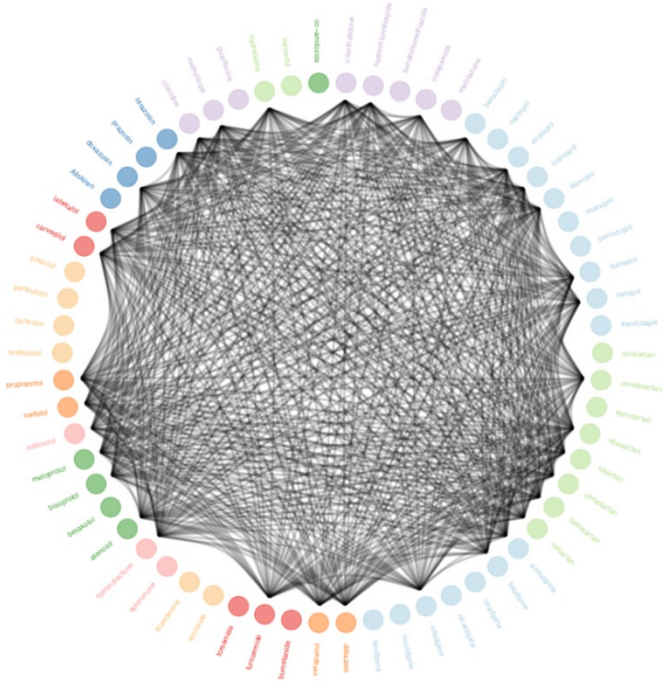
LEGEND

LARGE-SCALE EVIDENCE GENERATION AND EVALUATION IN A NETWORK OF DATABASES

LEGEND knowledge base for hypertension

Head-to-head HTN drug comparisons

- Trials: 40
- $N = 102 - [1148] - 33K$



- Comparisons: $10,278$
- $N = 3502 - [212K] - 1.9M$

LEGEND-HTN: Thiazide is better than ACEi for first-line anti-hypertension treatment

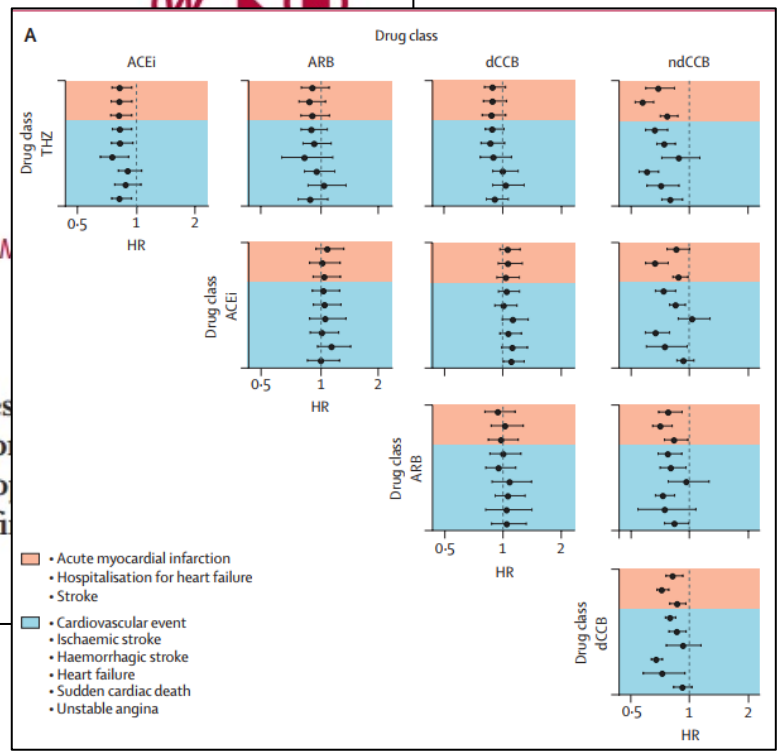
Articles

Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis

Marc A Suchard, Martijn J Schuemie, Harlan M Krumholz, Seng Chan You, RuiJun Chen, Nicole Pratt, Christian G Reich, Jon Duke, David M
George Hripcsak, Patrick B Ryan

Summary

Background Uncertainty remains about the optimal monotherapy for hypertension, with current guidelines
mending any primary agent among the first-line drug classes thiazide or thiazide-like diuretics, angiotensin-co
enzyme inhibitors, angiotensin receptor blockers, dihydropyridine calcium channel blockers, and non-dihydro
calcium channel blockers, in the absence of comorbid indications. Randomised trials have not further refi
choice.



Suchard, et al., Lancet 2019

JAMA | Original Investigation

Association of Ticagrelor vs Clopidogrel With Net Adverse Clinical Events in Patients With Acute Coronary Syndrome Undergoing Percutaneous Coronary Intervention

Seng Chan You, MD, MS; Yeunsook Rho, PhD; Behnood Bikdeli, MD, MS; Jiwoo Kim, MS; Anastasios Siapos, MSc; James Weaver, MSc; Ajit Londhe, MPH; Jaehyeong Cho, BS; Jimyung Park, BS; Martijn Schuemie, PhD; Marc A. Suchard, MD, PhD; David Madigan, PhD; George Hripcsak, MD, MS; Aakriti Gupta, MD, MS; Christian G. Reich, MD; Patrick B. Ryan, PhD; Rae Woong Park, MD, PhD; Harlan M. Krumholz, MD, SM

IMPORTANCE Current guidelines recommend ticagrelor as the preferred P2Y12 platelet inhibitor for patients with acute coronary syndrome (ACS), primarily based on a single large randomized clinical trial. The benefits and risks associated with ticagrelor vs clopidogrel in routine practice merits attention.

OBJECTIVE To determine the association of ticagrelor vs clopidogrel with ischemic and hemorrhagic events in patients undergoing percutaneous coronary intervention (PCI) for ACS in clinical practice.

Editorial page 1

JAMA Patient Page page 1

Audio and Supplemental content

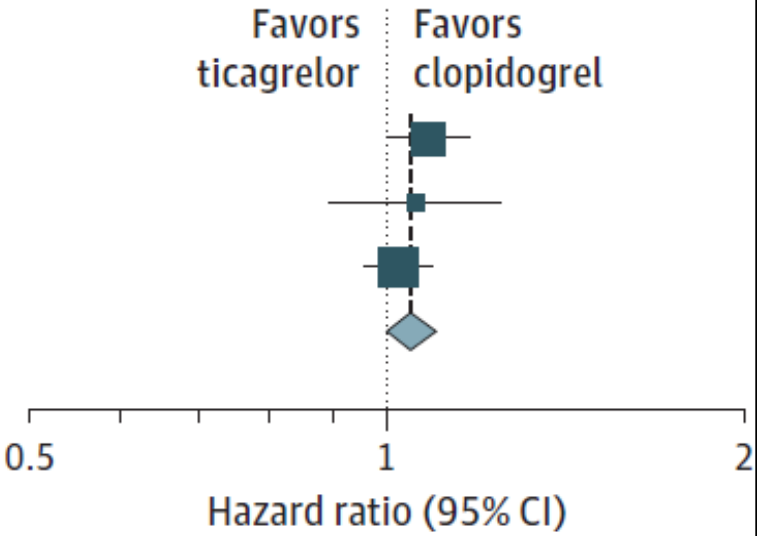
CME Quiz at jamacmelookup.com and CME Questions page 0

https://data.ohdsi.org/TicagrelorVsClopidogrel/

https://github.com/ohdsi-studies/ticagrelorVsClopidogrel/

Risk of the Primary Outcome (NACE) at 1 Year

Source	No. of events/total No. Ticagrelor	Clopidogrel	Hazard ratio (95% CI)
Optum electronic health record	1307/16 414	1192/16 414	1.08 (1.00-1.17)
IQVIA hospital	294/3998	272/3998	1.06 (0.90-1.24)
Health Insurance Review and Assessment	1883/10 878	1826/10 878	1.02 (0.96-1.09)
Overall: $I^2 = 0.0\%$; $P = .06$	3484/31 290	3290/31 290	1.05 (1.00-1.10)



inhibitor for patients with acute coronary syndrome (ACS), primarily based on a single large randomized clinical trial. The benefits and risks associated with ticagrelor vs clopidogrel in routine practice merits attention.

OBJECTIVE To determine the association of ticagrelor vs clopidogrel with ischemic and hemorrhagic events in patients undergoing percutaneous coronary intervention (PCI) for ACS in clinical practice.

JAMA Patient Page page 1

Audio and Supplemental content

CME Quiz at jamacmelookup.com and CME Questions page 0

https://data.ohdsi.org/TicagrelorVsClopidogrel/

https://github.com/ohdsi-studies/ticagrelorVsClopidogrel/

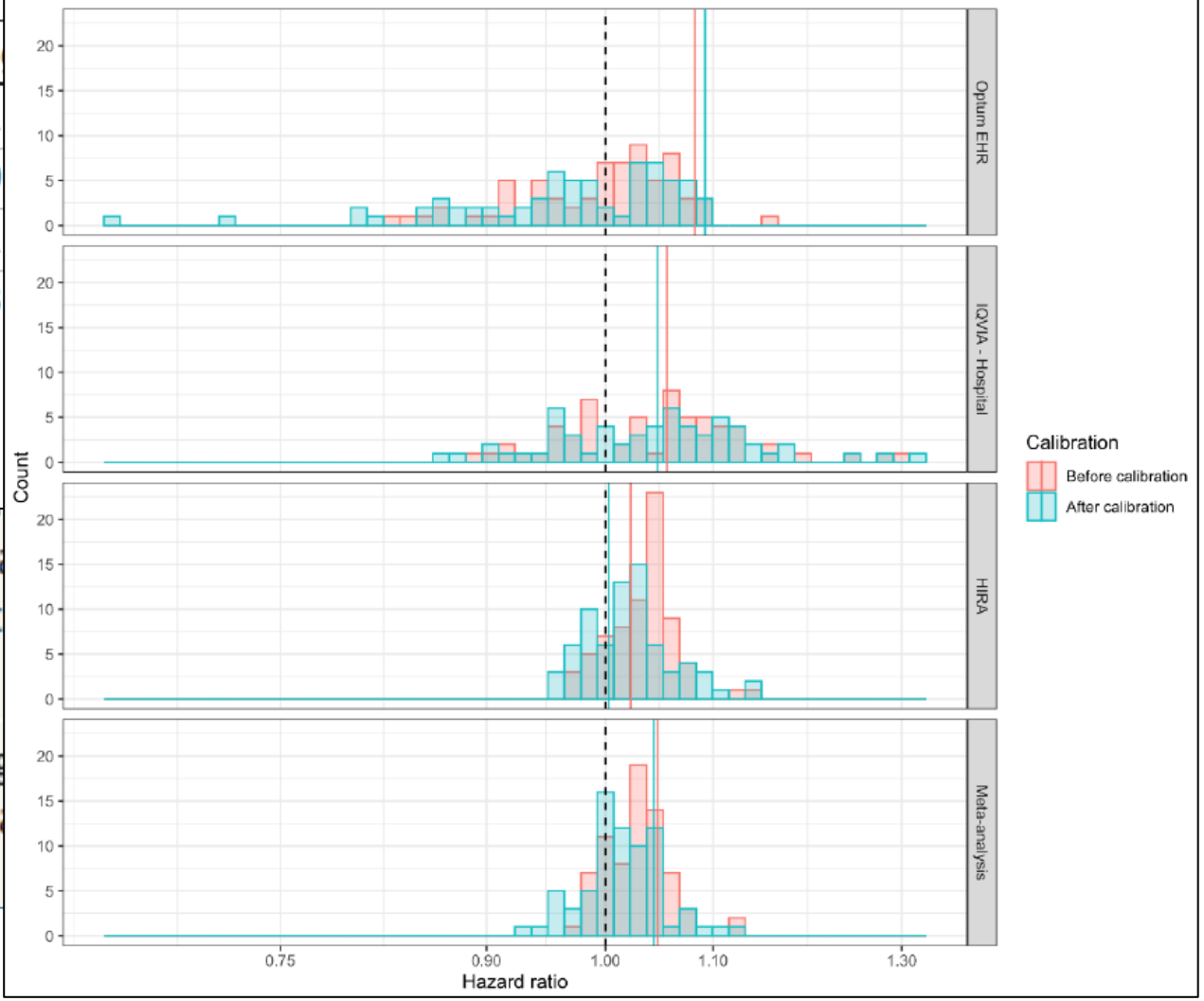
Risk of the Primary Outcome (NACE) at 1 Year

Source	No. of events/total No. Ticagrelor	Clopido
Optum electronic health record	1307/16 414	1192/1
IQVIA hospital	294/3998	272/39
Health Insurance Review and Assessment	1883/10 878	1826/1
Overall: $I^2 = 0.0\%$; $P = .06$	3484/31 290	3290/3

inhibitor for patients with acute coronary syndrome (ACS), prima
randomized clinical trial. The benefits and risks associated with t
routine practice merits attention.

OBJECTIVE To determine the association of ticagrelor vs clopidog
hemorrhagic events in patients undergoing percutaneous coron
in clinical practice.

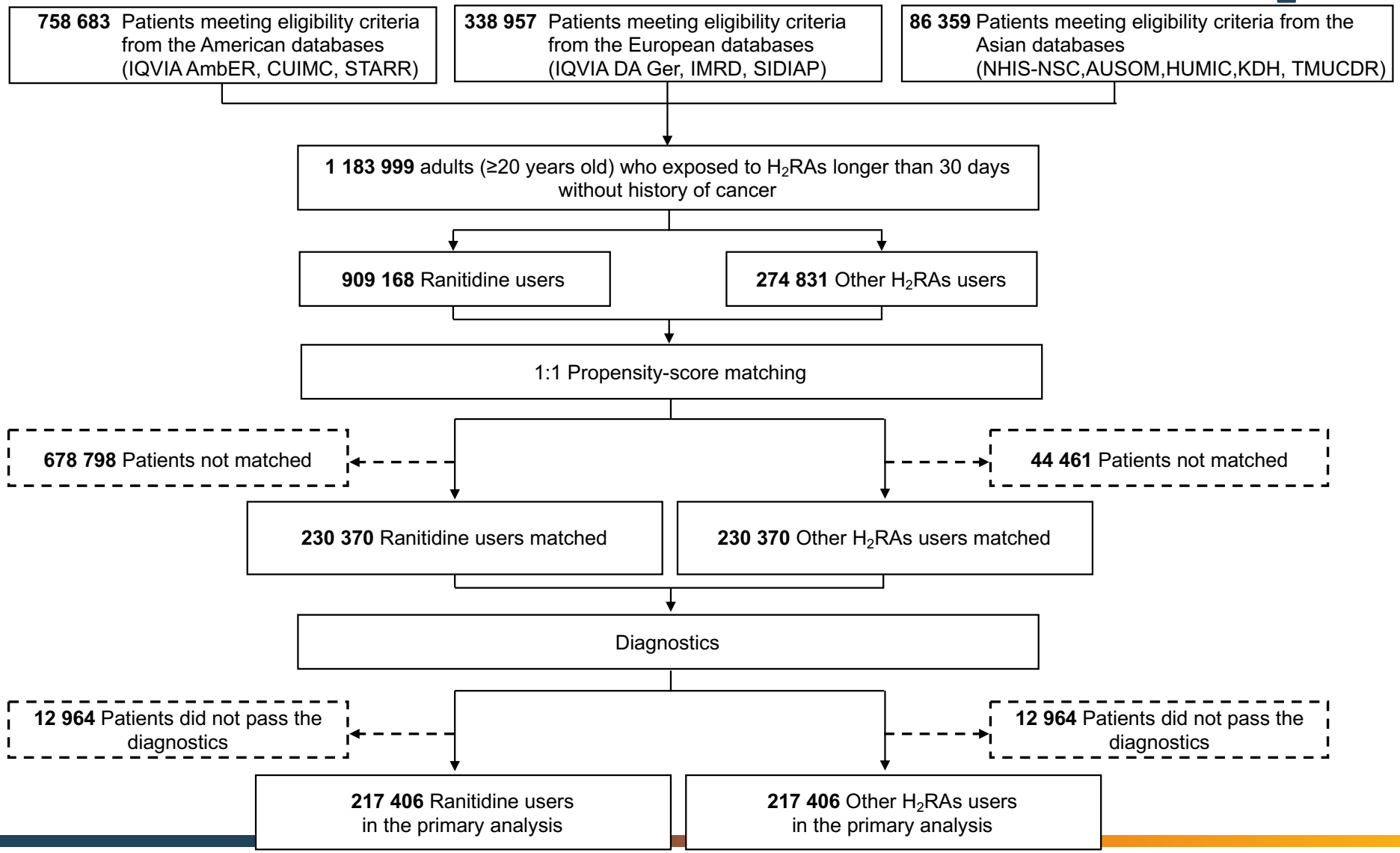
eFigure 7. Distribution of risk estimates for NACE from 144 analyses before and after empirical calibration



https://data.ohdsi.org/TicagrelorVsClopidogrel/

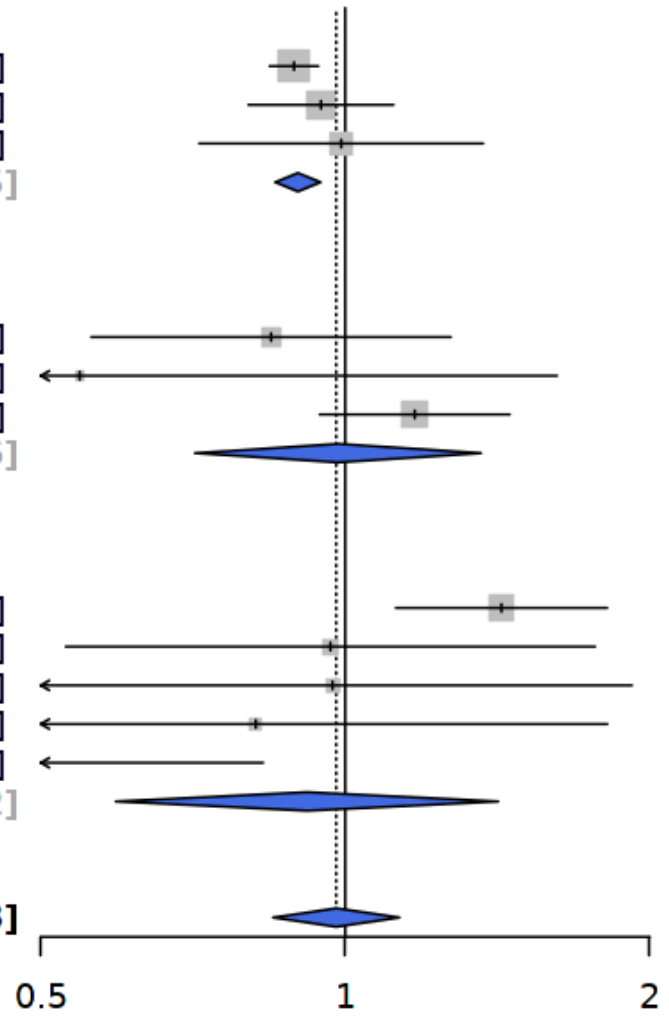
https://github.com/ohdsi-studies/ticagrelorVsClopidogrel/

Ongoing international study: The risk of cancer in ranitidine vs other H_2RAs



The risk of cancer in ranitidine vs other H_2RAs

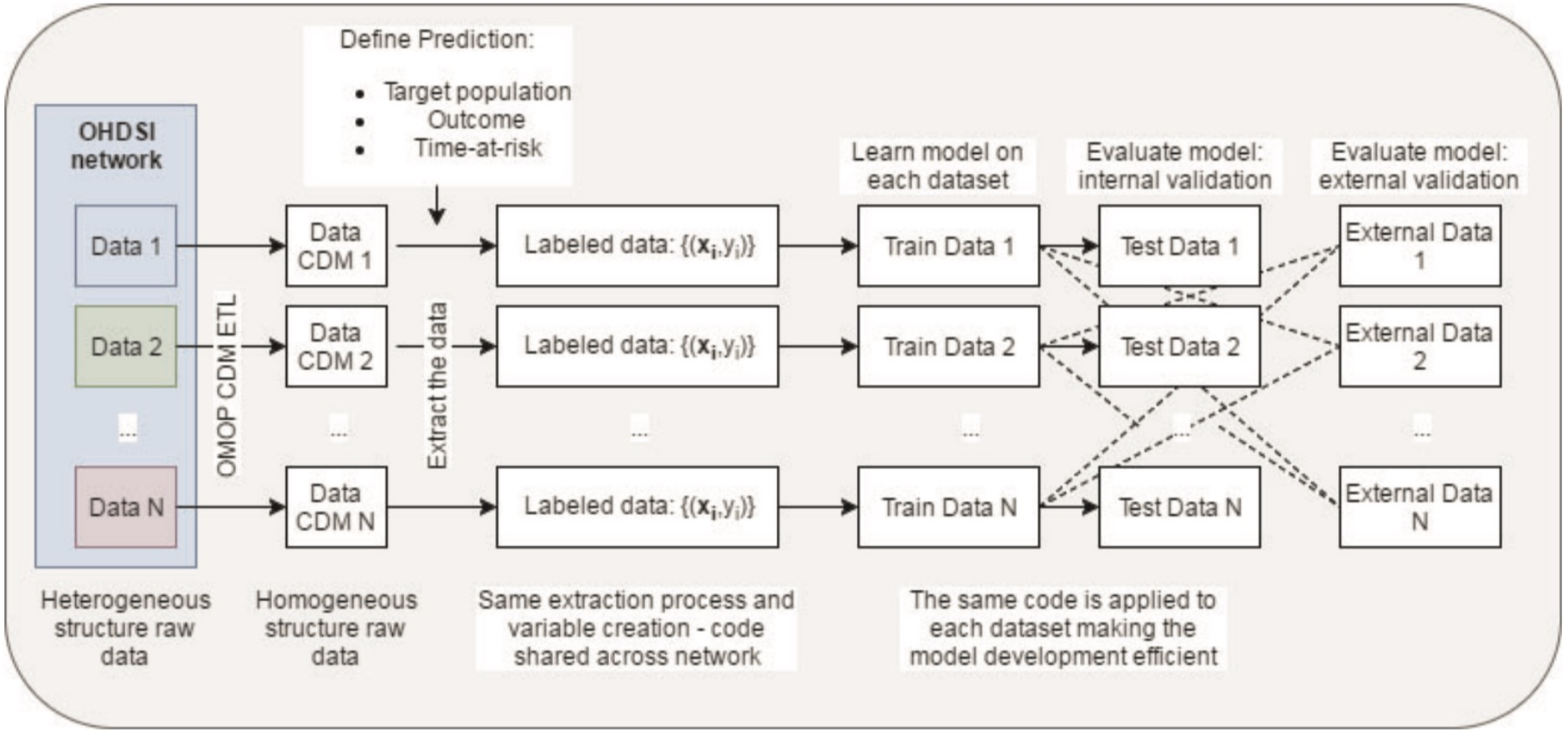
Source	Ranitidine	Other H2 blocker	HR	95% CI
America				
IQVIA Ambulatory EMR	2380/248009	2643/248009	0.89	[0.84; 0.94]
CUIMC	276/12738	290/12738	0.95	[0.80; 1.12]
STARROMOP	73/3604	75/3604	0.99	[0.72; 1.37]
Overall			**0.90**	**[0.85; 0.95]**
Heterogeneity: I^2 = 0%				
Europe				
IQVIA DA Germany	43/3546	50/3546	0.84	[0.56; 1.27]
imrd1903	5/697	11/697	0.55	[0.18; 1.62]
SIDIAP	176/4241	155/4241	1.17	[0.94; 1.45]
Overall			**0.98**	**[0.71; 1.36]**
Heterogeneity: I^2 = 41.7%				
ASIA				
NHIS-NSC	165/14118	112/14118	1.43	[1.12; 1.82]
AUSOM	21/2300	22/2300	0.97	[0.53; 1.77]
HUMIC	18/1883	16/1883	0.97	[0.49; 1.92]
KDH	11/1155	14/1155	0.82	[0.37; 1.82]
TMUCDR	7/1973	20/1973	0.35	[0.14; 0.83]
Overall			**0.92**	**[0.59; 1.42]**
Heterogeneity: I^2 = 65.3%				
Overall			**0.98**	**[0.85; 1.13]**
Heterogeneity: I^2 = 60.5%				
Test for subgroup differences: χ^2_2 = 0.30, df = 2 (p = 0.86)				



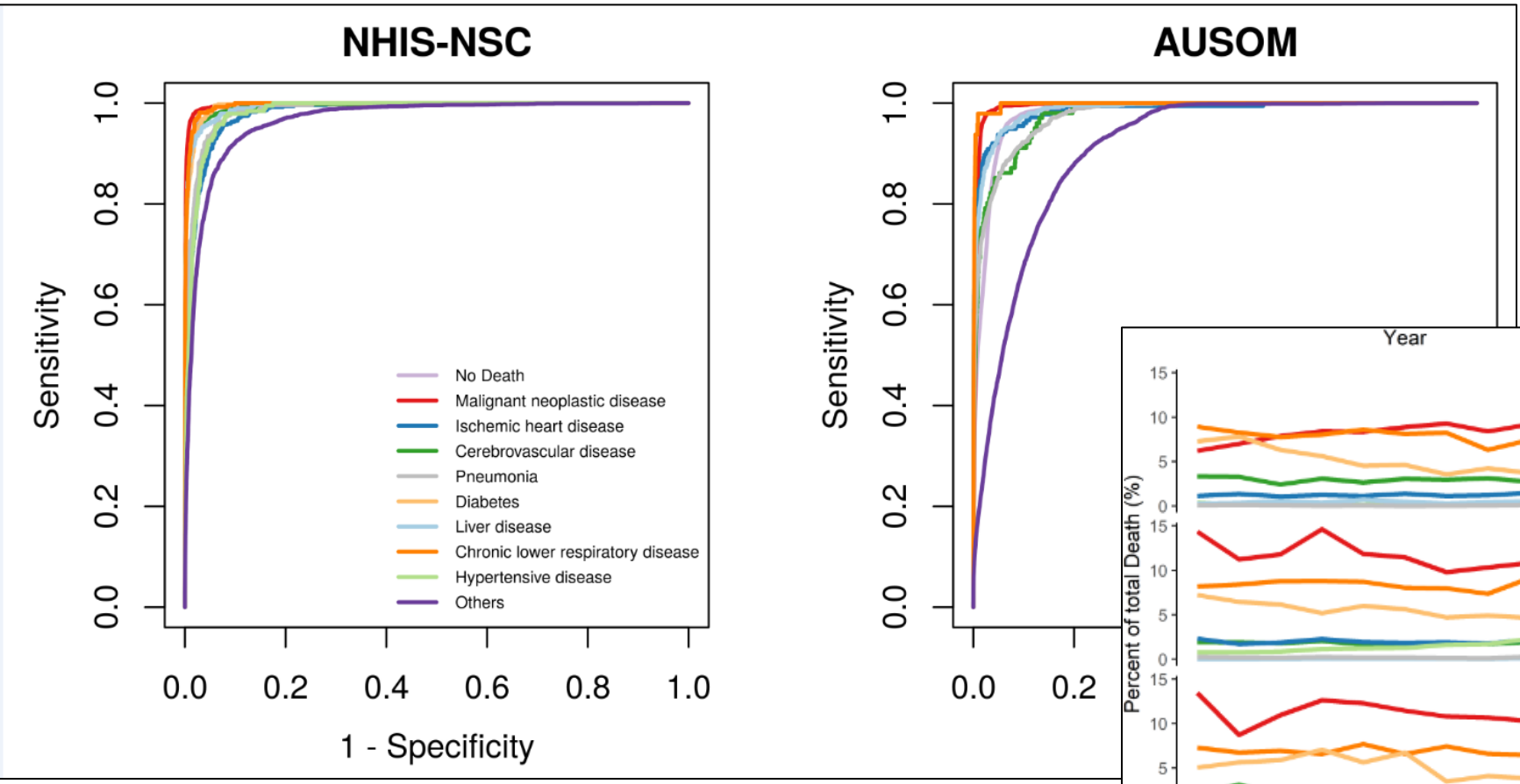
Contents

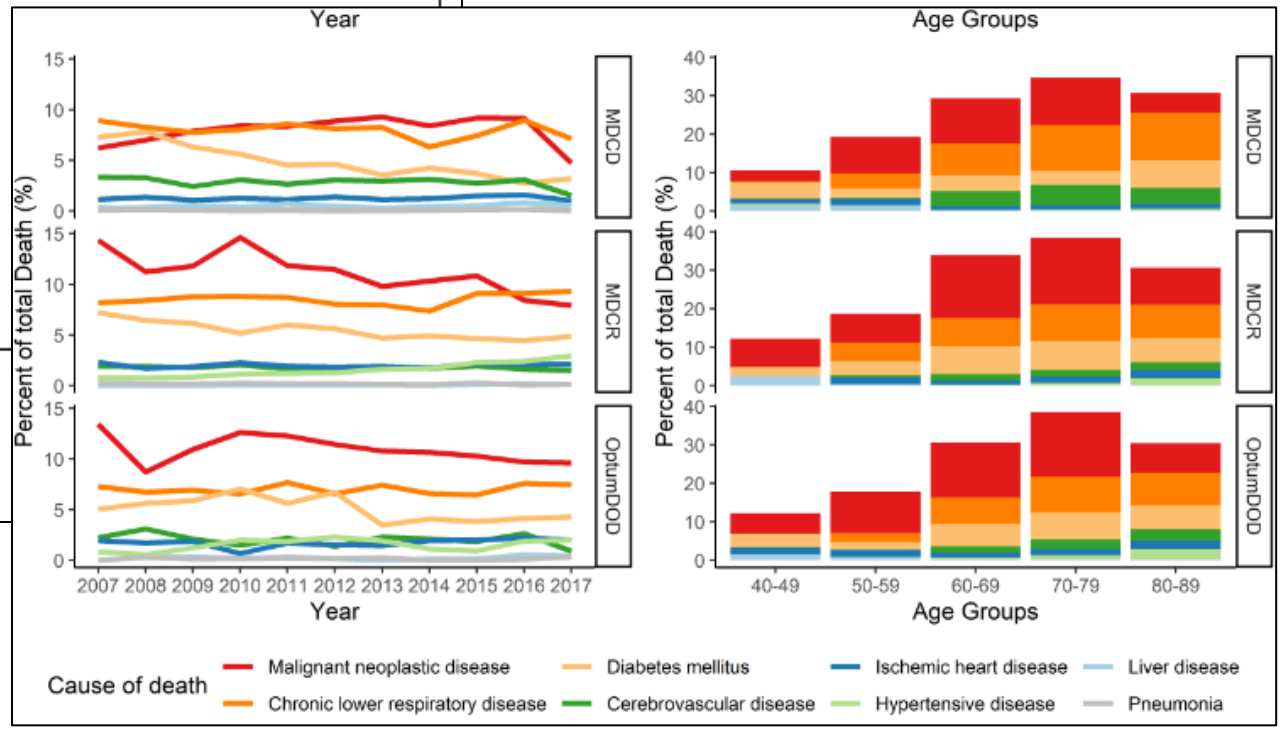
- Common Data Model (CDM) / OHDSI
- Open-source and Characterization
- Population-level estimation: Large-scale Evidence Generation and Evaluation in a Network of Databases
- Patient-Level Prediction

Bring the algorithm to the data, not data to the algorithm

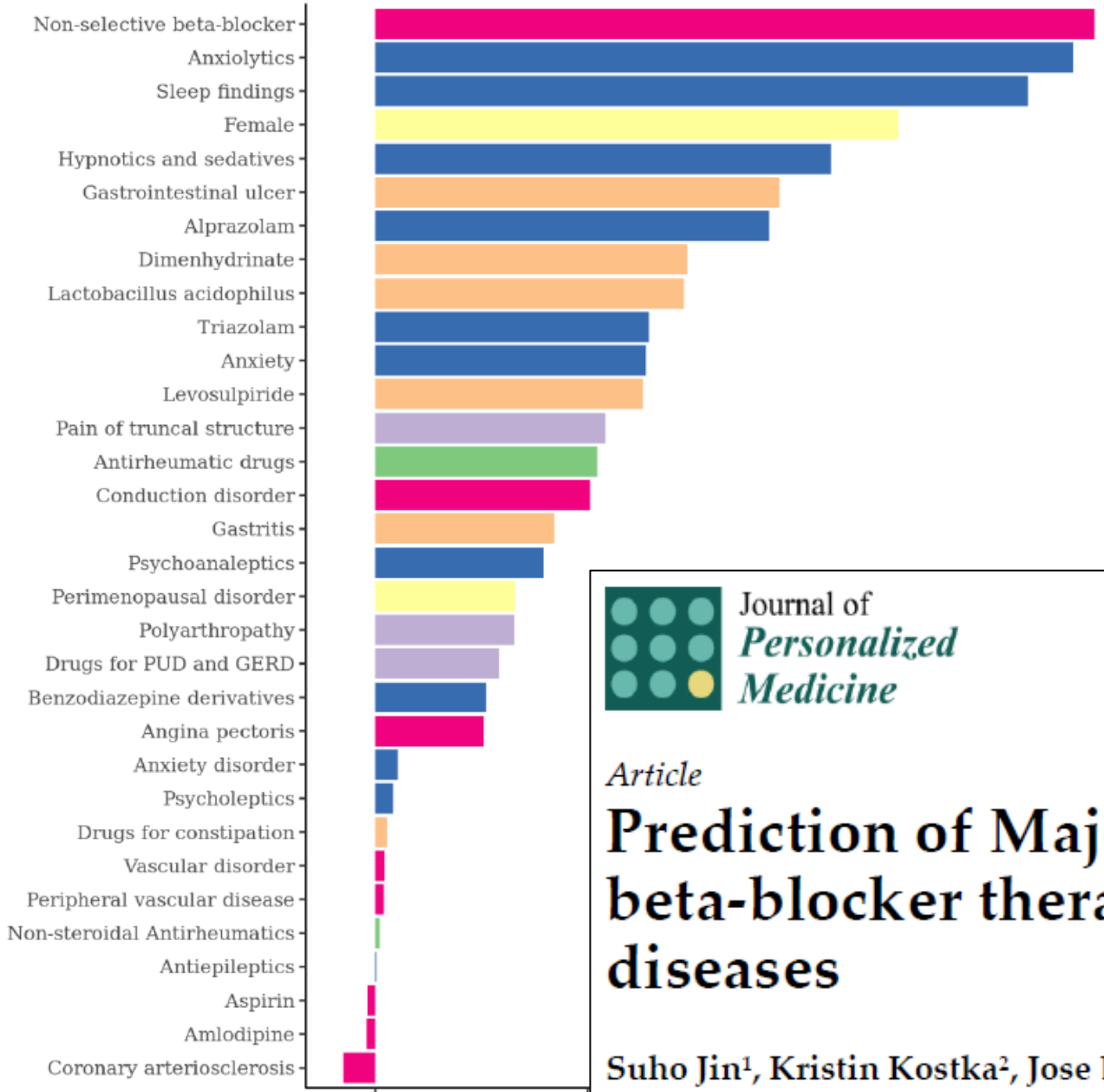


Robust machine-learning model to predict the cause of death based on distributed research network / CDM





Ckim, SCYou, et al., JAMIA (accepted)



Journal of *Personalized Medicine*

MDPI

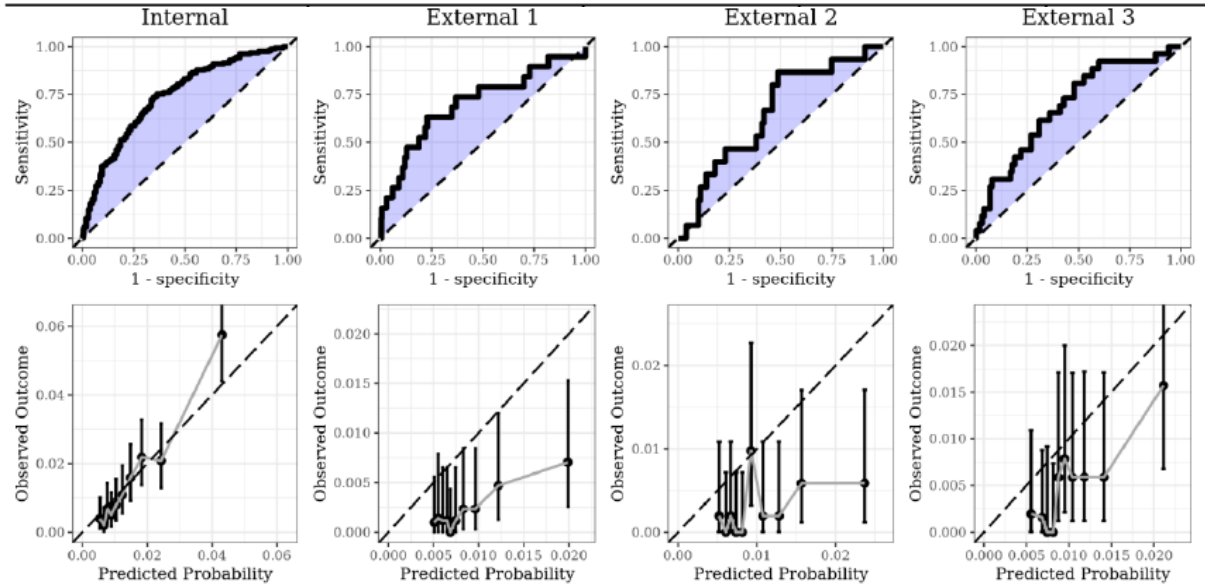
Article

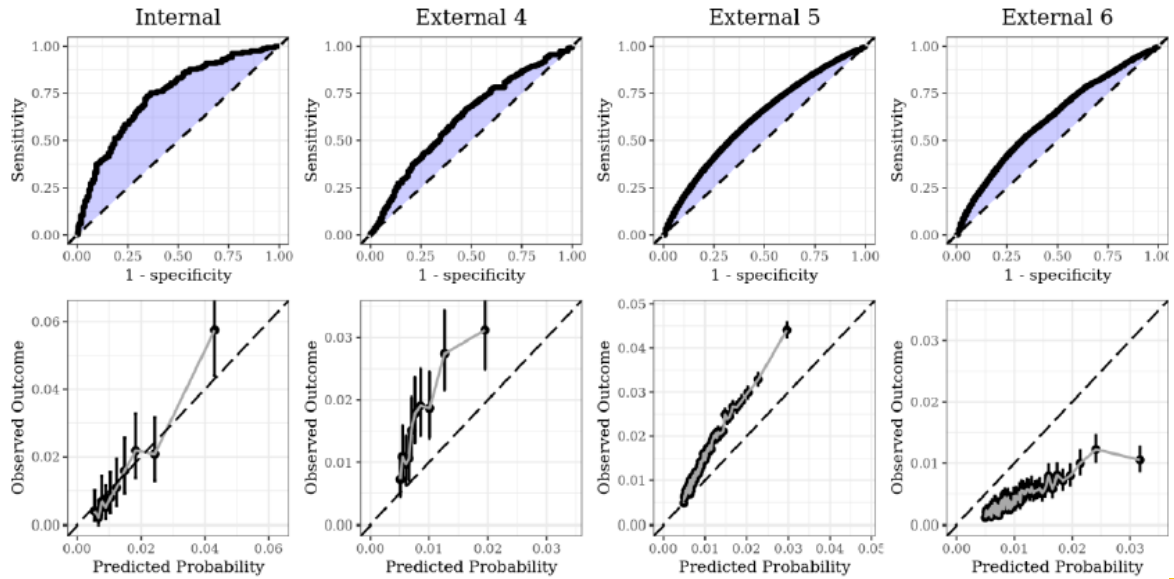
Prediction of Major Depressive Disorder following beta-blocker therapy in patients with cardiovascular diseases

Suho Jin[1], Kristin Kostka[2], Jose D. Posada[3], Yeesuk Kim[4], Seung In Seo[5], Dong Yun Lee[6], Nigam H. Shah[7], Sungwon Roh[8], Young-Hyo Lim[9], Sun Geu Chae[10], Uram Jin[11], Sang Joon Son[12], Christian Reigh[13], Peter R. Rijnbeek[14], Rae Woong Park[15,*], and Seng Chan You[16,*]

Table 2. Performance of the model in internal and external validations

Validation set	Name	n	Outcome	Incidence (%)	AUC	Sensitivity	Specificity
Internal	NHIS	10078	154	1.53	0.74	83.1%	49.5%
External 1	Ajou	8511	19	0.22	0.71	78.9%	49.0%
External 2	Hanyang	5112	15	0.29	0.66	86.7%	49.4%
External 3	Kandong	5097	26	0.51	0.70	80.8%	49.9%
External 4	STARR	26,258	439	1.67	0.62	77.2%	40.4%
External 5	OpenClaims	4,295,013	59,045	1.38	0.62	75.1%	40.2%
External 6	AmbEMR	883,198	3,342	0.38	0.62	75.4%	40.1%





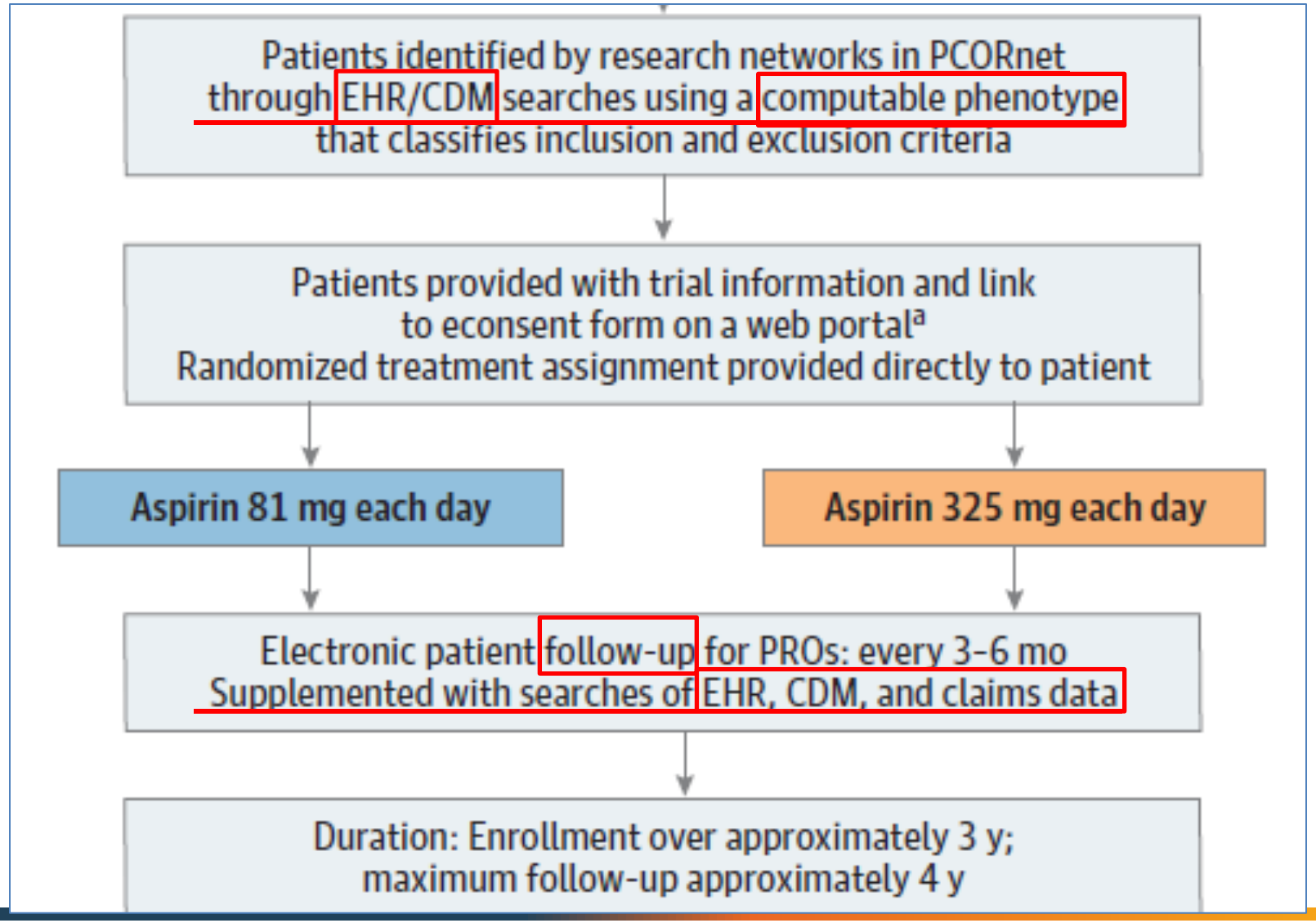
JAMA Cardiology | Special Communication

Rationale and Design of the Aspirin Dosing—A Patient-Centric Trial Assessing Benefits and Long-term Effectiveness (ADAPTABLE) Trial

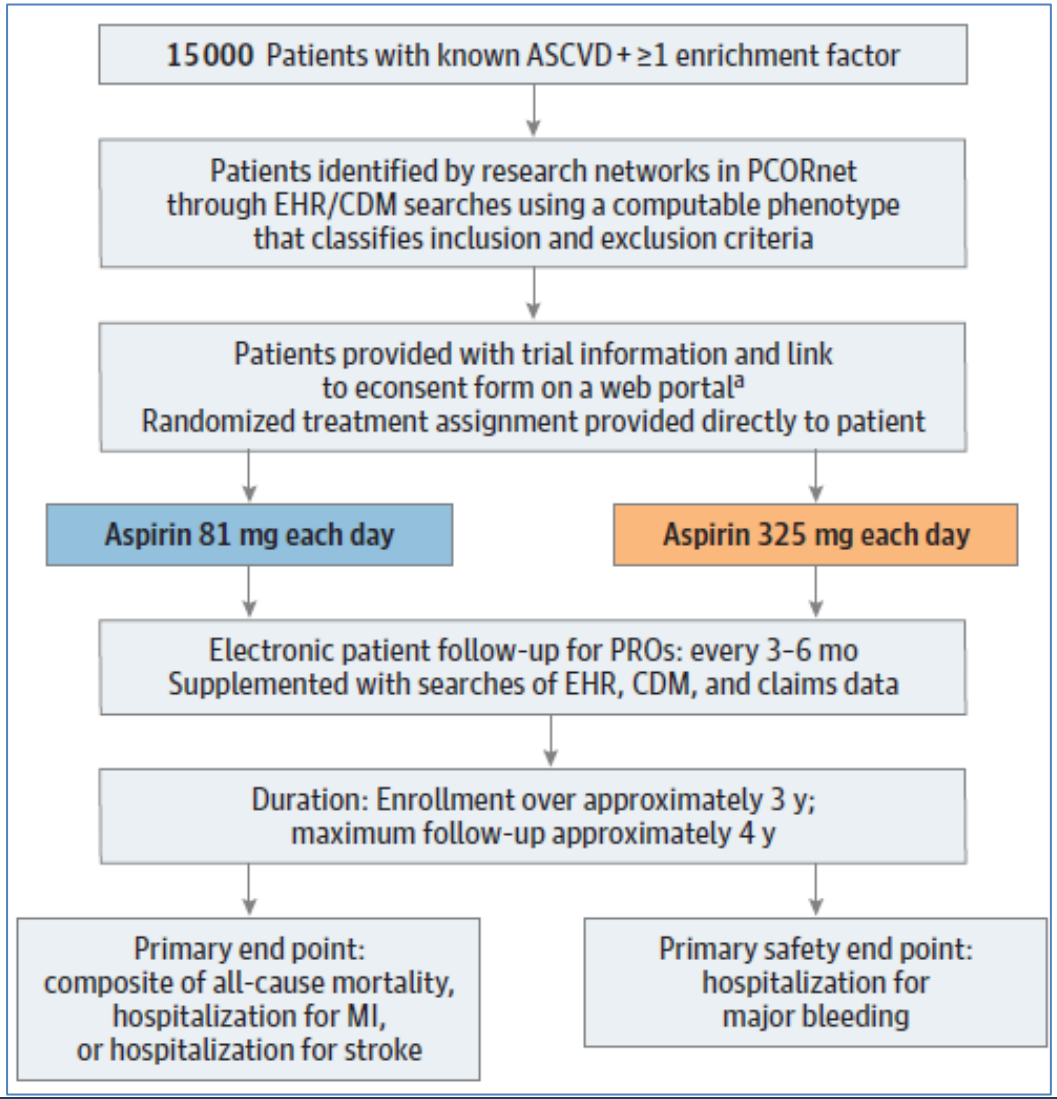
DESIGN, SETTING, AND PARTICIPANTS This pragmatic, open-label, patient-centered, randomized clinical trial is being conducted in 15 000 patients within the National Patient-Centered Clinical Research Network (PCORnet), a distributed research network of partners including clinical research networks, health plan research networks, and patient-powered research networks across the United States. Patients with established ASCVD treated in routine clinical practice within the network are eligible. Patient recruitment began in April 2016. Enrollment was completed in June 2019. Final follow-up is expected to be completed by June 2020.

MAIN OUTCOMES AND MEASURES The primary efficacy end point is the composite of all-cause mortality, hospitalization for nonfatal myocardial infarction, or hospitalization for a nonfatal stroke. The primary safety end point is hospitalization for major bleeding associated with a blood-product transfusion. End points are captured through regular queries of the health systems' common data model within the structure of PCORnet's distributed data environment.

Pragmatic **C**linical **T**rial based on **Nationwide CDM** data network

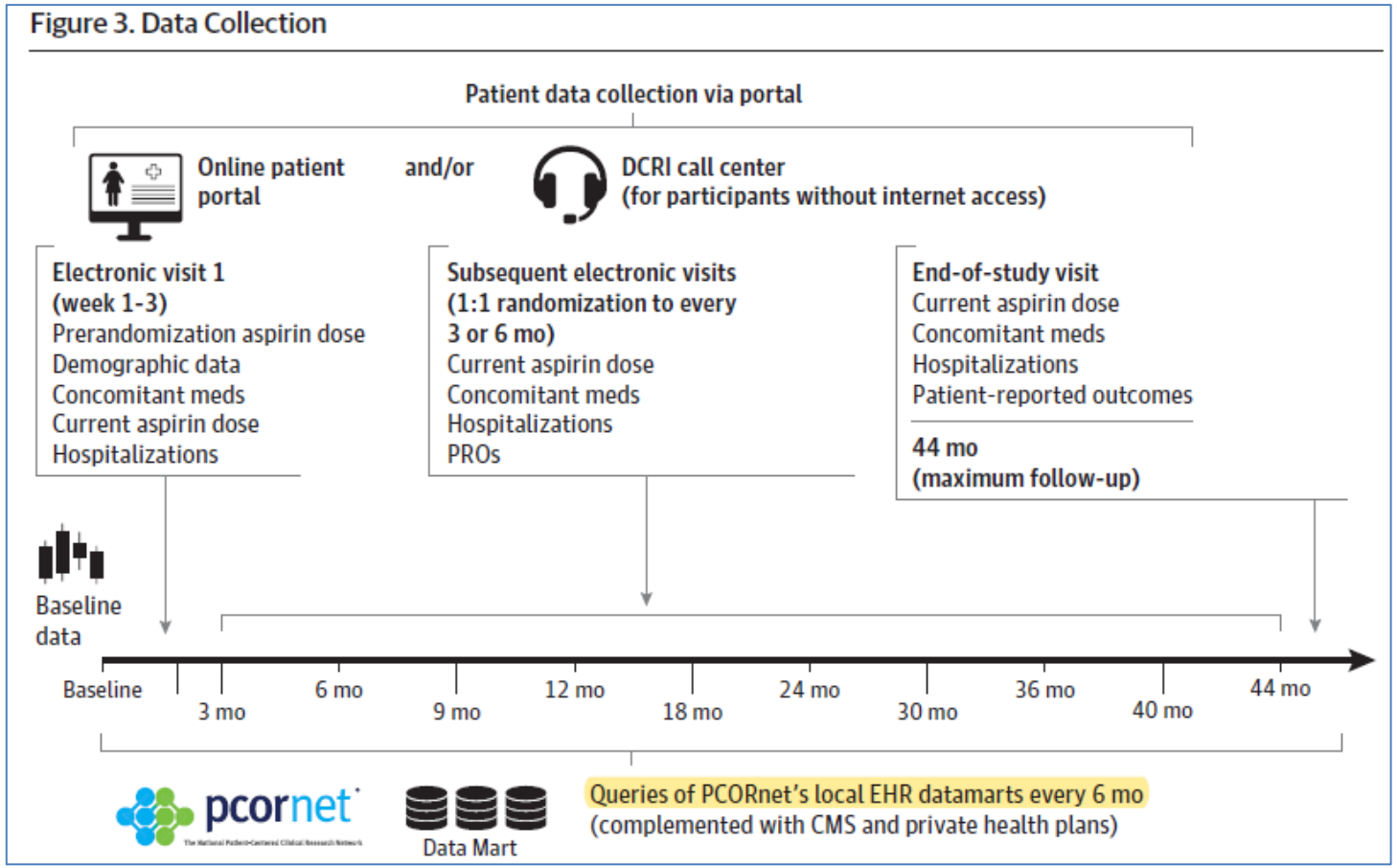


Pragmatic Clinical Trial based on Nationwide CDM data network



Annals of Internal Medicine IDEAS AND OPINIONS

The ADAPTABLE Trial and PCORnet: Shining Light on a New Research Paradigm



Marquis-Gravel *et al., JAMA Cardiology, 2020*

Hernandez *et al., Annals of Internal Medicine, 2015*

Longitudinal Expansion of Data: Integration with Standardized Nationwide Claim Data

심사평가원, '보건의료 분야 결합전문기관' 최초 지정

- 조직·인력, 시설·시스템, 개인정보 안전 및 보안 등 지정요건 충족 -

□ 건강보험심사평가원(원장 김선민, 이하 '심사평가원')이 보건복지부에서 주관한 '데이터 결합전문기관 지정을 위한 지정심사위원회'를 거쳐, 10월 29일「보건의료 분야 결합전문기관」으로 최초 지정되었다.

○ 결합전문기관 지정은 데이터 이용 활성화를 통한 신산업 육성을 위해 가명정보*의 결합이 가능토록 개정된 데이터3법('20.8.5.시행)에 따라, 가명정보에 대한 결합전문기관을 지정하는 것으로 올해 최초 도입됐다.

사전규격 상세 (용역)

[공공기관 요청규격서]

참조번호	계약부-3452호	사전규격등록번호	923681
품명(사업명)	공통데이터모델(CDM) 표준용어 매핑사전 검증 및 데이터 변환		
배정예산액	₩ 395,000,000		
공개일시	2020/10/30 14:52	의견등록마감일시	2020/11/02 23:5
공고기관	건강보험심사평가원 편정훈(033-739-2612)		
수요기관	건강보험심사평가원		
SW사업대상여부	비대상	납품(완수)기한 (납품일수)	계약 후 90일 이내

http://www.g2b.go.kr/pt/menu/selectSubFrame.do?framesrc=/pt/menu/frameTgong.do?url=http://www.g2b.go.kr:8101/ep/tbid/tbidList.do?taskClCds=&bidNm=&searchDtType=1&fromBidDt=2020/10/30&toBidDt=2020/11/06&fromOpenBidDt=&toOpenBidDt=&radOrgan=1&instNm=%BD%C9%BB%E7%C6%F2%B0%A1%BF%F8&area=&regYn=Y&bidSearchType=1&searchType=1

Longitudinal Expansion of Data: Integration with Standardized Nationwide Claim Data

UPDATED. 2020-11-06 15:24 (금)

MEDICAL Observer

무작위 임상연구+리얼월드=세계 임상연구 '1위' 가능?

박선혜 기자 | 승인 2018.10.26 06:15 | 댓글 0

> 실용적(Pragmatic) RCT 연구 중요성 대두...등록사업·의료기록 등 이용해 RCT 진행
> 국내 전문가 "우리나라 실용적 RCT 할 수 있는 최적 조건 갖췄다"

국내 학계에서는 실용적 RCT가 우리나라에 최적화된 연구 디자인이라고 평가한다. 우리나라는 전 국민 단일 건강보험 체계로, 국민건강보험공단, 건강보험심사평가원에 국민들의 의료 데이터가 모이기 때문이다.

박 교수는 "우리나라는 건보공단 또는 심평원 자료를 통해 환자가 복용한 치료제, 입원 여부 등에 대한 정보를 모두 알 수 있다. 전 국민 의료기록이 모니터링된다"며 "때문에 환자가 병원에 방문할 때마다 일일이 예후를 확인하지 않아도 된다. 실용적 RCT를 잘할 수 있는 국가적인 시스템이 마련돼 있다. 이를 잘 활용하면 우리나라가 전 세계 임상연구 1등을 할 수 있다"고 피력했다.

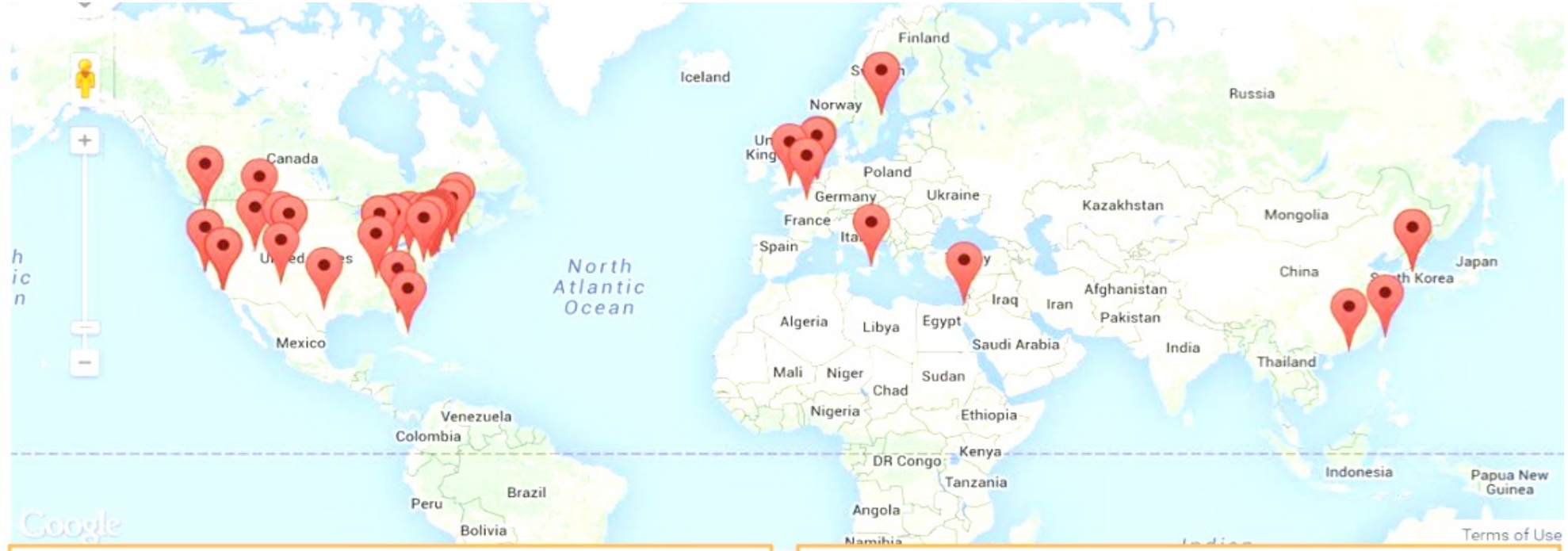
김 교수는 "대규모 등록사업이나 코호트가 잘 구축돼 있으면 실용적 RCT를 시작하기 쉽다"며 "우리나라는 단일 건강보험 체계라는 점에서 다른 나라에 비해 실용적 RCT를 진행하기 훨씬 유리한 상황"이라고 강조했다.

'개인정보보호법'에 발목 잡혀..."좋은 시스템이 있는데 시작하지 못한다"

하지만 이러한 장점에도 불구하고 우리나라에서는 실용적 RCT를 하기 위한 걸음마조차 떼지 못한 실정이다. 가장 큰 걸림돌은 '개인정보' 문제다. 개인정보보호법이 강화되고 개인의료정보가 민감한 개인정보로 여겨져 환자들의 연구 참여 동의를 받기란 쉽지 않기 때문이다.

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:

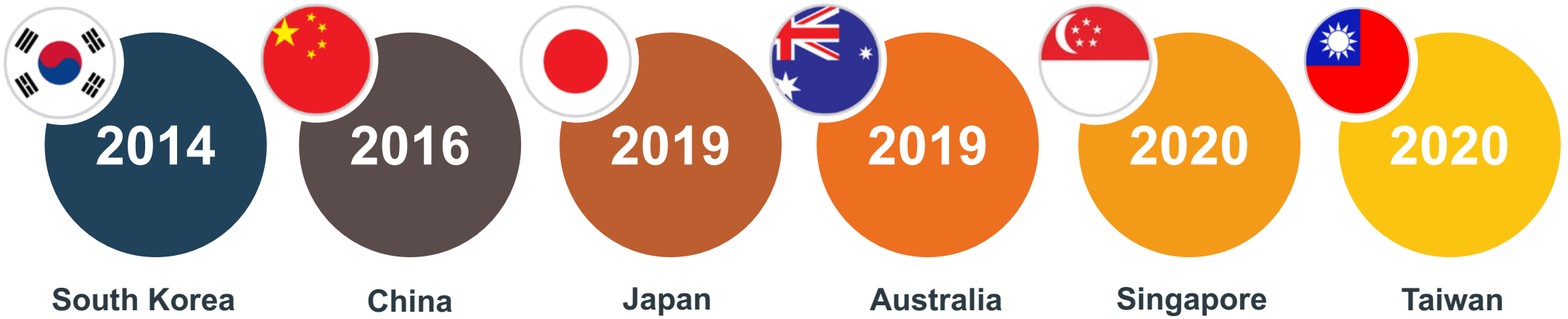
- >100 researchers in academia, industry and government
- >10 countries

OHDSI Data Network:

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

https://www.ohdsi.org/

OHDSI APAC Chapters Introduction



The Book of OHDSI

The Book of OHDSI Korea

OHDSI-Korea

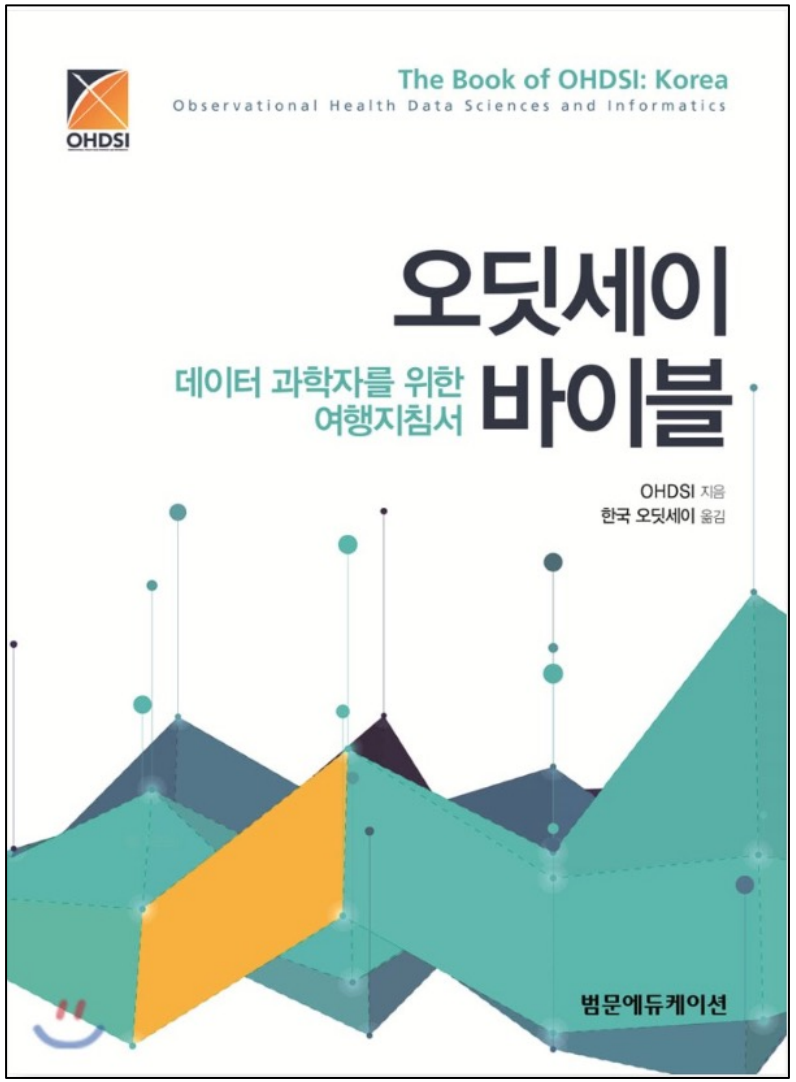
2019-10-18

서문

이 책은 관찰 보건 데이터 과학 및 정보학 (Observational Health Data Sicence and Informatics : OHDSI) 공동작업에 대한 내용을 담고있다. OHDSI 커뮤니티에 의해 작성된 이 책은 OHDSI 관련 모든 지식의 중앙저장소 역활을 담당하고자 쓰여졌으며 오픈소스 개발 도구들을 통해 커뮤니티에 의해 관리되는 생명력있는 문서로 계속 진화하고 있다. 또한 ohdsi-korea.github.io/TheBookOfOhdsiKorea/에서 온라인으로 항상 최신 버전의 책을 무료로 받아 볼 수 있으며 실물로 구입을 원할 경우 Amazon 등에서 구입이 가능하다.

이 책의 목표

이 책은 OHDSI 관련 모든 지식의 중앙저장소 역할을 담당하고자 쓰여졌으며 OHDSI 커뮤니티, OHDSI 데이타 기준과 OHDSI 도구들에 중점을 두었다. OHDSI의 초보자와 숙련자 모두를 위해 현실적으로 필요 이론과 사용방법에 대한 교육을 제공하는 실용적인 부분에 목표를 두고 있다. 이 책을 읽은 뒤 당신은 OHDSI란 무엇인가, 또한 그 여정에 어떻게 동참할 것인가에 관하여 이해하게 될 것이다. 또한 공통 데이터 모델(CDM)과 표준화된 용어들이 무엇인지, 이러한 것들이 관찰 보건 데이터베이스의 표준화에 어떻게 사용되는지 알게 될 것이다. 이 데이터에 대해 Clinical characterization, Population-level estimation,




Thank You
for your time