Observational Health Data Sciences and Informatics, Interoperability, and Research

George Hripcsak, MD, MS
Columbia University Irving Medical Center
NewYork-Presbyterian Hospital
Observational Health Data Sciences and Informatics

Observational Health Data Sciences and Informatics

(OHDSI, as “Odyssey”)

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

A multi-stakeholder, interdisciplinary, international collaborative with a coordinating center at Columbia University

OHDSI's global research community

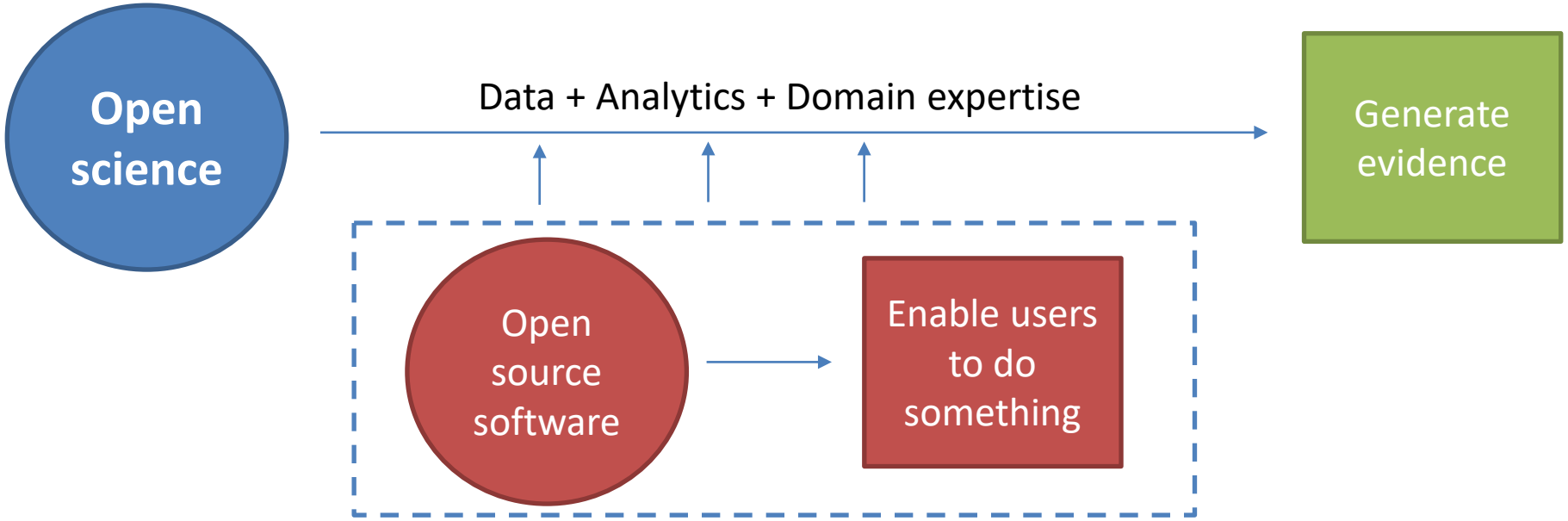
- >300 collaborators from 30 different countries
- Experts in informatics, statistics, epidemiology, clinical sciences
- Active participation from academia, government, industry, providers
- Records on about 600 million unique patients in >100 databases

http://ohdsi.org/who-we-are/collaborators/

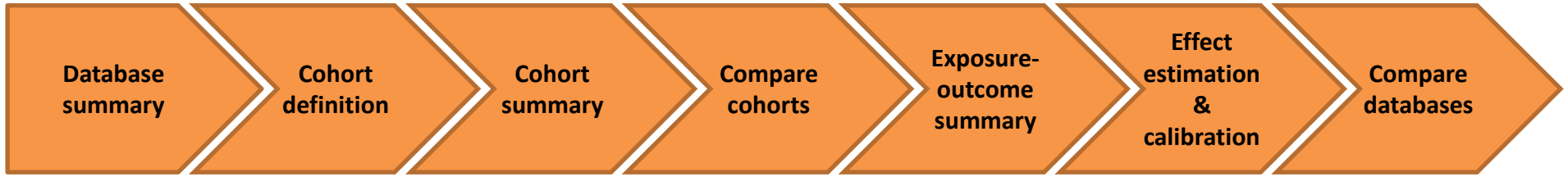
Evidence OHDSI seeks to generate from observational data

- **Clinical characterization - tally**
  - Natural history: Who has diabetes, and who takes metformin?
  - Quality improvement: What proportion of patients with diabetes experience complications?
- **Population-level estimation - cause**
  - Safety surveillance: Does metformin cause lactic acidosis?
  - Comparative effectiveness: Does metformin cause lactic acidosis more than glyburide?
- **Patient-level prediction - predict**
  - Precision medicine: Given everything you know about me, if I take metformin, what is the chance I will get lactic acidosis?
  - Disease interception: Given everything you know about me, what is the chance I will develop diabetes?

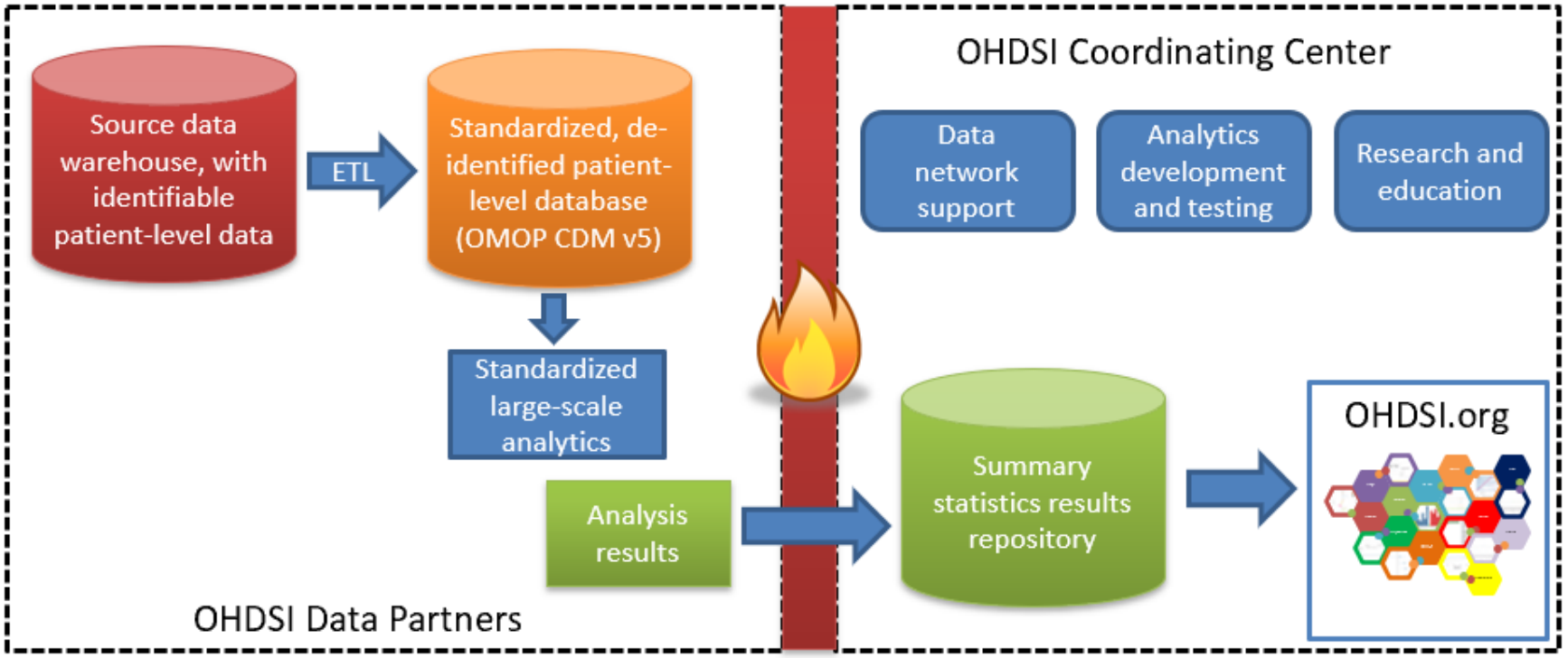
Open Science



Standardized, transparent workflows

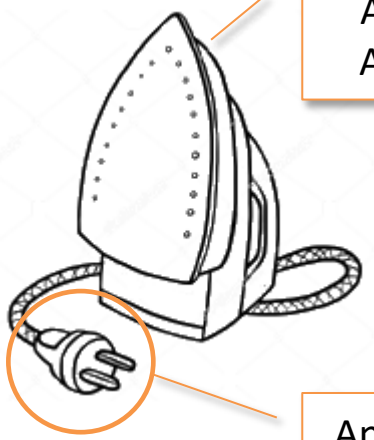


How OHDSI Works



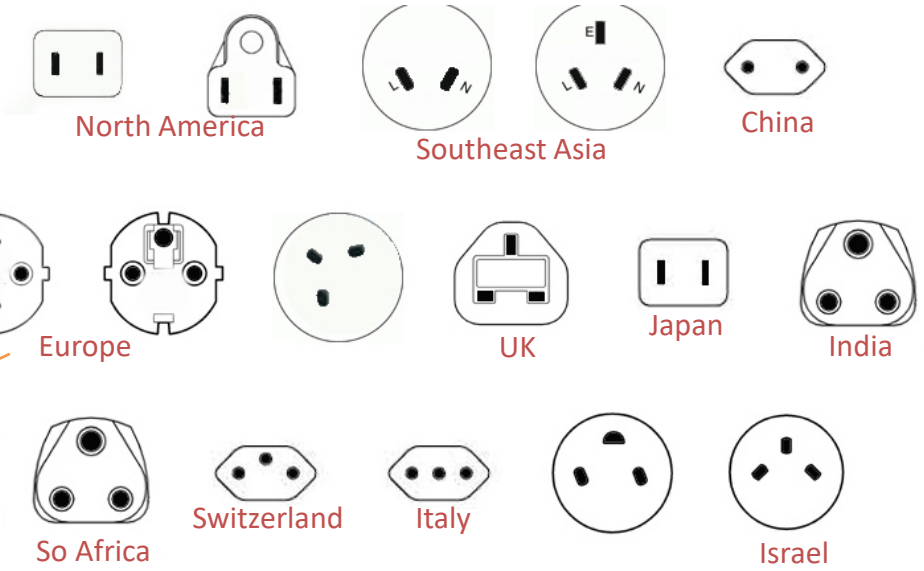
Current Approach: “One Study – One Script“

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

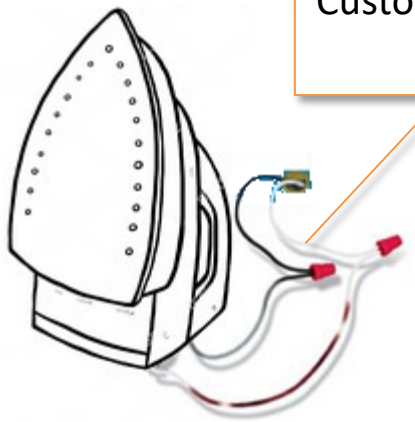


Analytical method: Adherence to Drug

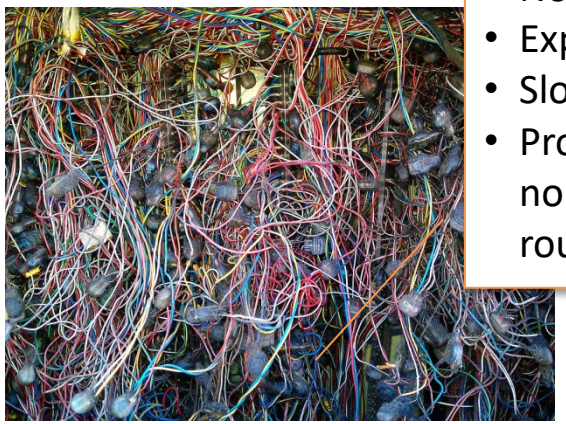
Application to data



Current solution:

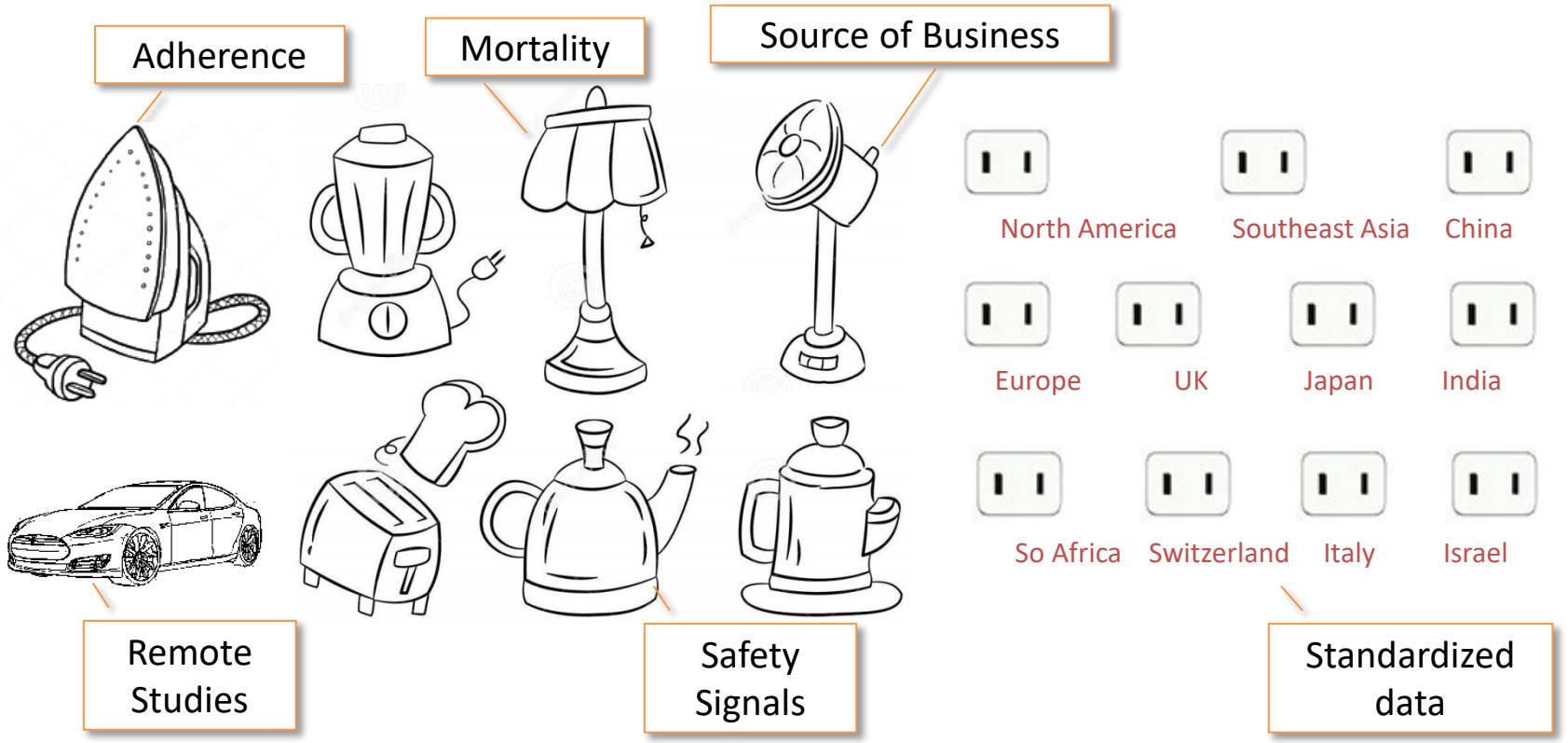


Custom script for each study



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

Solution: Standardized Data and Analytics

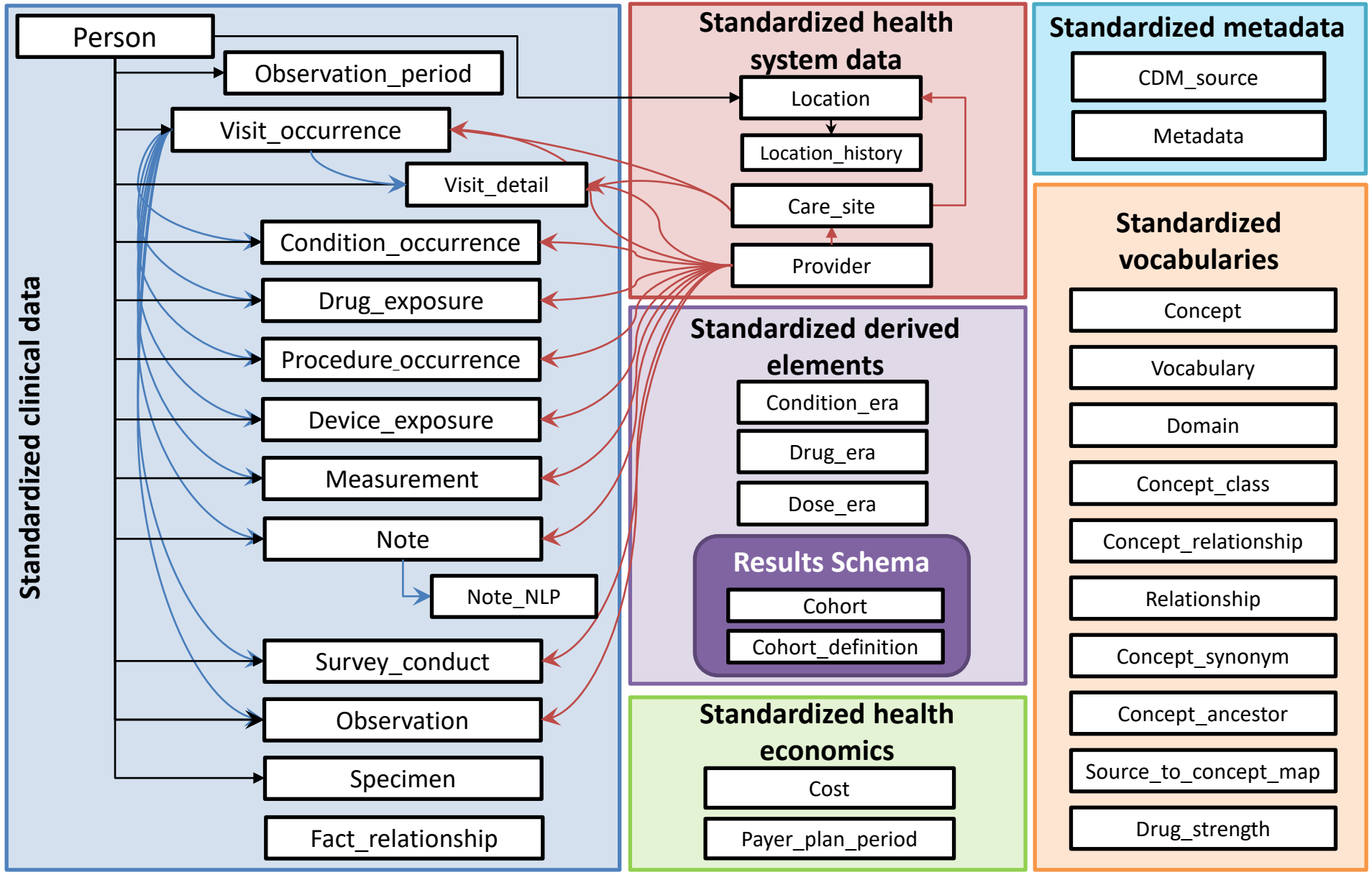


1. ATLAS, Remote Studies
   - Standard Cohorts
   - Standardized Analytics
2. OMOP CDM
   - Standardized Format
   - Standardized Coding

Common Data Model

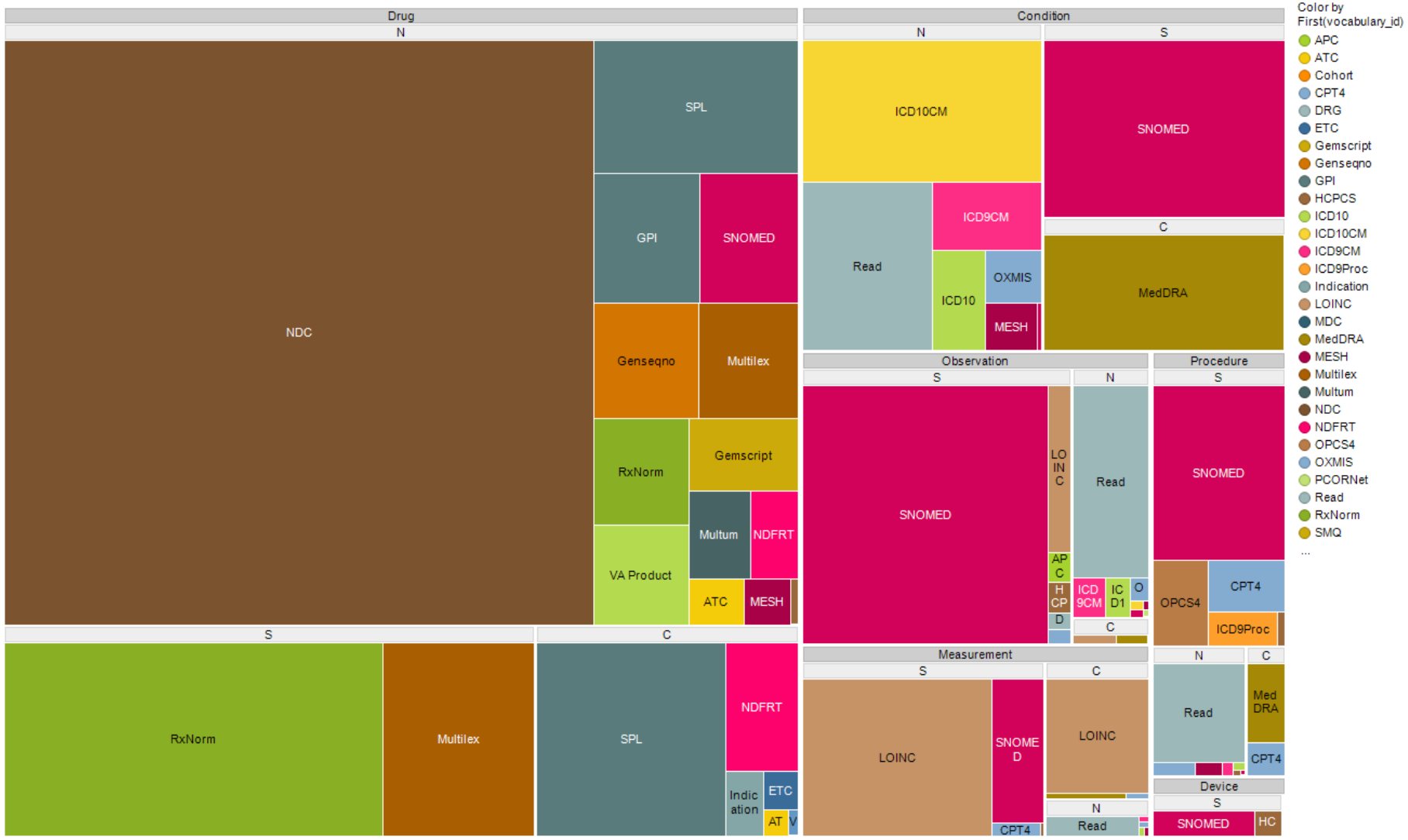
- OMOP
  - Observational Medical Outcomes Partnership
  - (Origin of OHDSI; kept the data model name)
- Components
  - **Schema** – tables where you put data
  - **Vocabulary** – what codes go in the table
  - **Conventions** – how to store data
- Open committee structure to govern it
  - Contracted vocabulary maintenance

Deep information model
OMOP CDM



Extensive vocabularies

Breakdown of OHDSI concepts by domain, standard class, and vocabulary



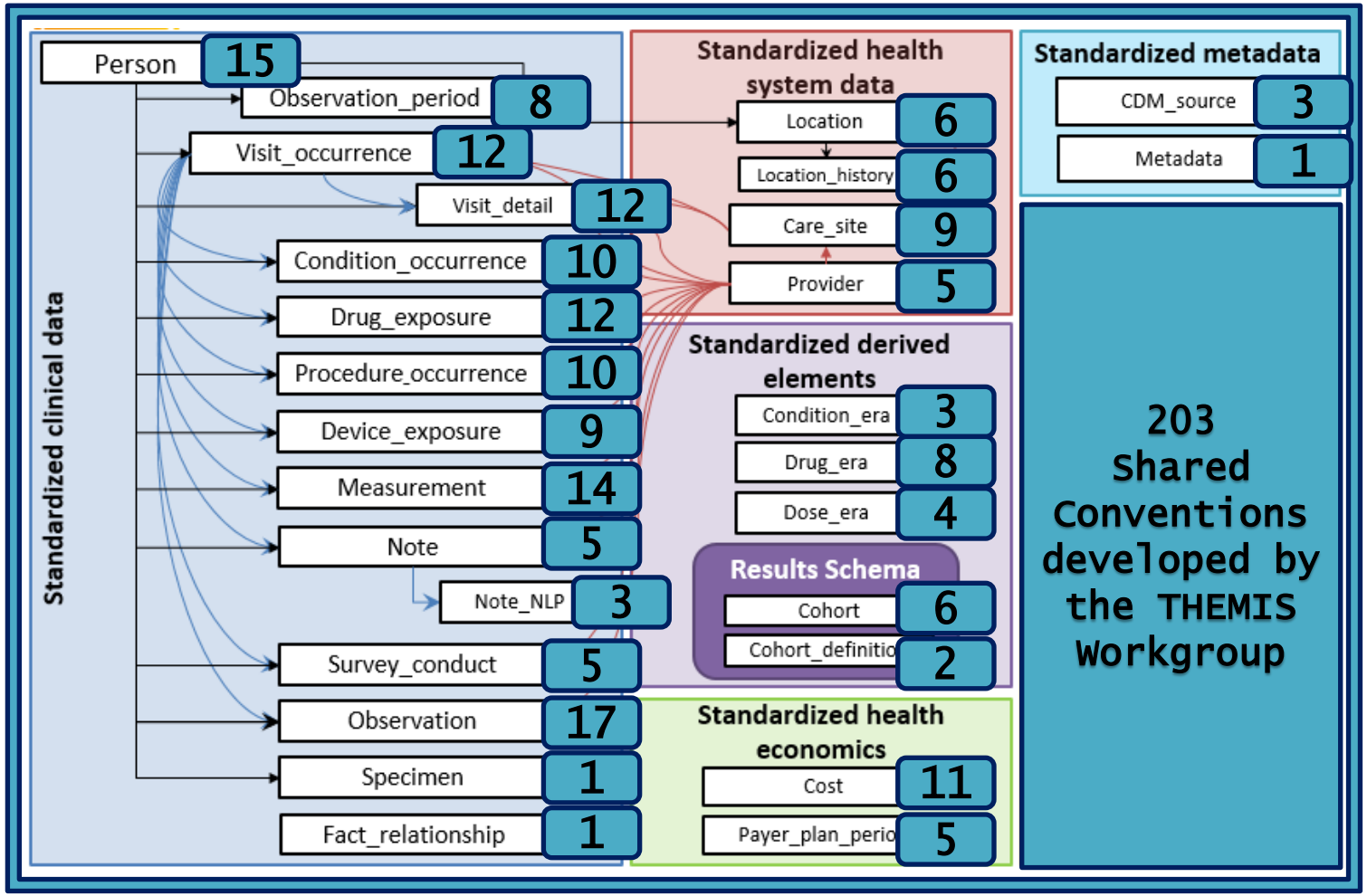
OHDSI's standardized vocabularies

- 153 Vocabularies across 41 domains
  - MU3 standards: SNOMED, RxNorm, LOINC
  - Disparate sources: ICD9CM, ICD10(CM), Read, NDC, Gemscript, CPT4, HCPCS…
- >9 million concepts
  - >3.3 million standard concepts
  - >5.1 million source codes
  - >629,000 classification concepts
- >55 million concept relationships
- >84 million ancestral relationships

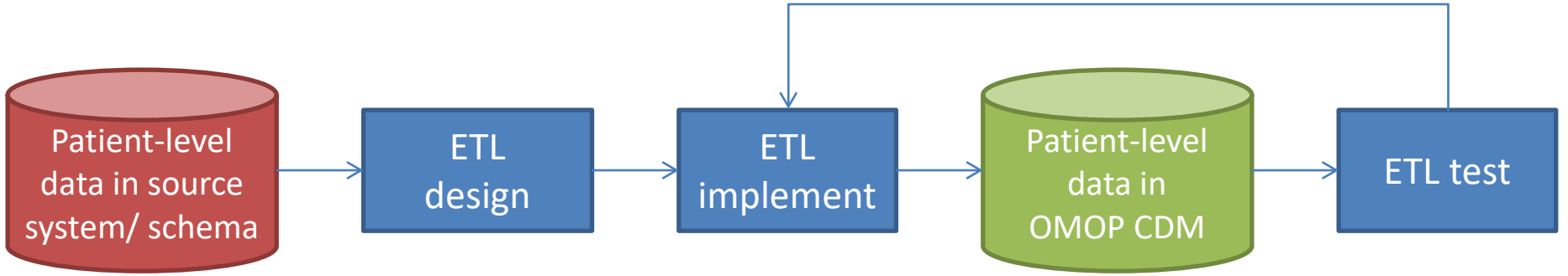
Standard vocabularies

- Bring the world's data to a core set
- RxNorm
  - RxNorm Extensions - cover non-US
- LOINC
- SNOMED CT
  - How to address non-US data
- In many areas, stuck adopting several
  - Procedures, …

Standardized conventions



Preparing your data



OHDSI tools built to help

WhiteRabbit: profile your source data

RabbitInAHat: map your source structure to CDM tables and fields

ATHENA: standardized vocabularies for all CDM domains

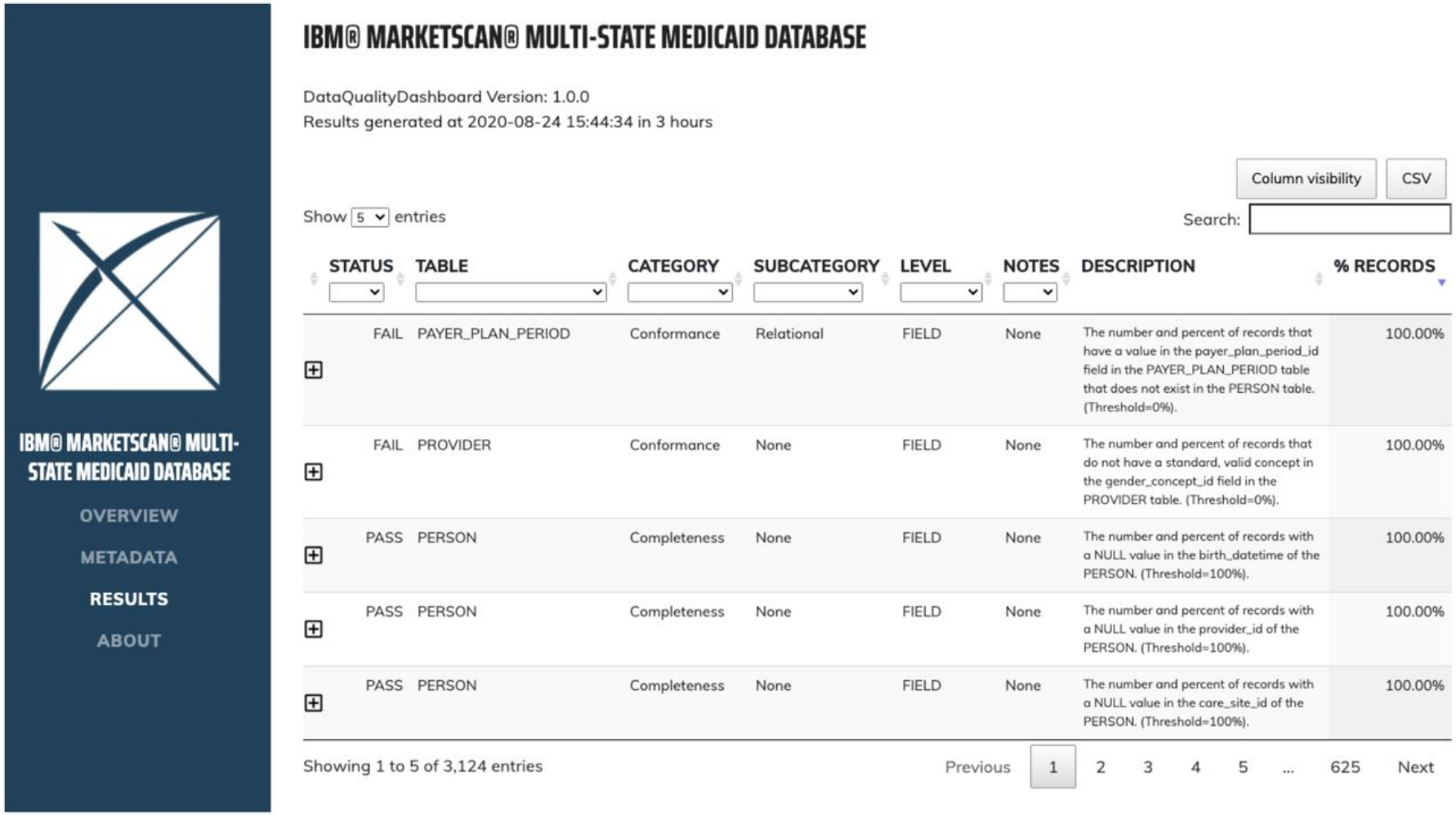
Usagi: map your source codes to CDM vocabulary

CDM: DDL, index, constraints for Oracle, SQL Server, PostgresQL; Vocabulary tables with loading scripts

ACHILLES: profile your CDM data; review data quality assessment; explore population-level summaries

OHDSI Forums:
Public discussions for OMOP CDM Implementers/developers

Data Quality Dashboard



IBM® MARKETSCAN® MULTI-STATE MEDICAID DATABASE

OVERVIEW

METADATA

RESULTS

ABOUT

IBM® MARKETSCAN® MULTI-STATE MEDICAID DATABASE

DataQualityDashboard Version: 1.0.0

Results generated at 2020-08-24 15:44:34 in 3 hours

Column visibility | CSV

Show 5 entries

Search:

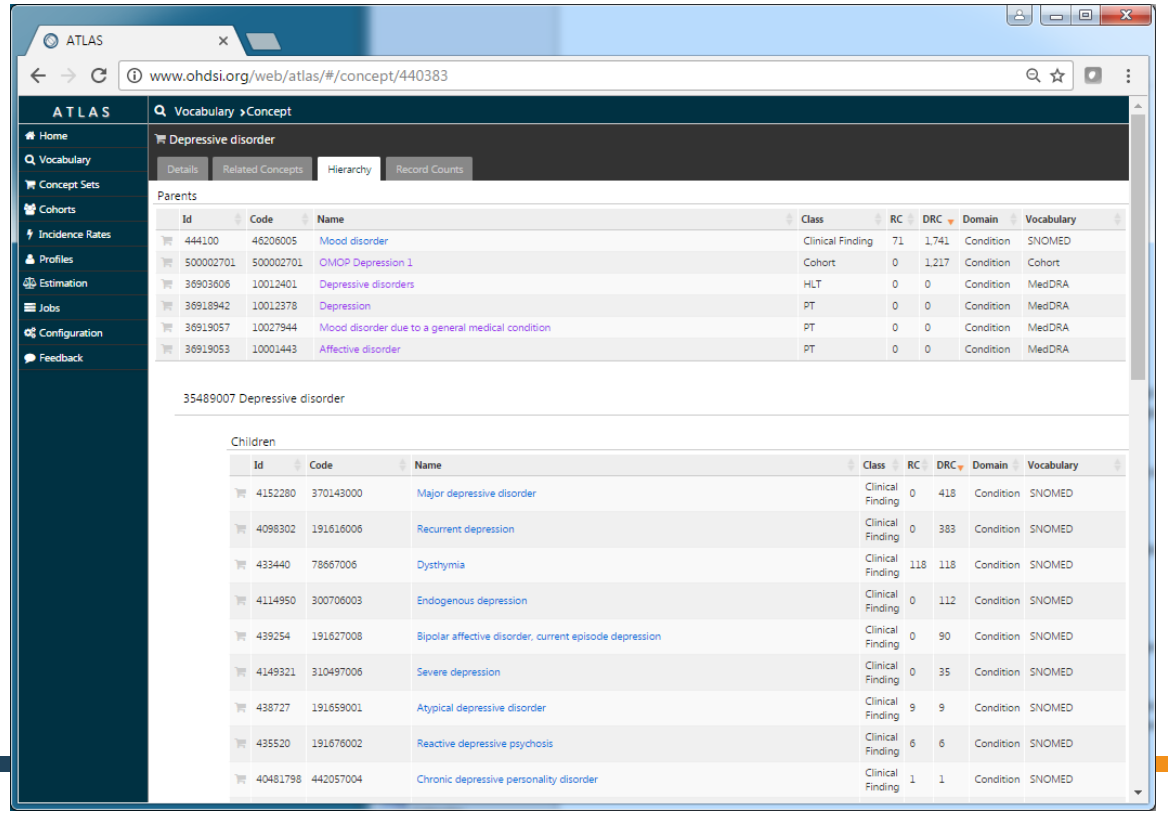
STATUS	TABLE	CATEGORY	SUBCATEGORY	LEVEL	NOTES	DESCRIPTION	% RECORDS
FAIL	PAYER_PLAN_PERIOD	Conformance	Relational	FIELD	None	The number and percent of records that have a value in the payer_plan_period_id field in the PAYER_PLAN_PERIOD table that does not exist in the PERSON table. (Threshold=0%).	100.00%
FAIL	PROVIDER	Conformance	None	FIELD	None	The number and percent of records that do not have a standard, valid concept in the gender_concept_id field in the PROVIDER table. (Threshold=0%).	100.00%
PASS	PERSON	Completeness	None	FIELD	None	The number and percent of records with a NULL value in the birth_datetime of the PERSON. (Threshold=100%).	100.00%
PASS	PERSON	Completeness	None	FIELD	None	The number and percent of records with a NULL value in the provider_id of the PERSON. (Threshold=100%).	100.00%
PASS	PERSON	Completeness	None	FIELD	None	The number and percent of records with a NULL value in the care_site_id of the PERSON. (Threshold=100%).	100.00%

Showing 1 to 5 of 3,124 entries

Previous 1 2 3 4 5 ... 625 Next

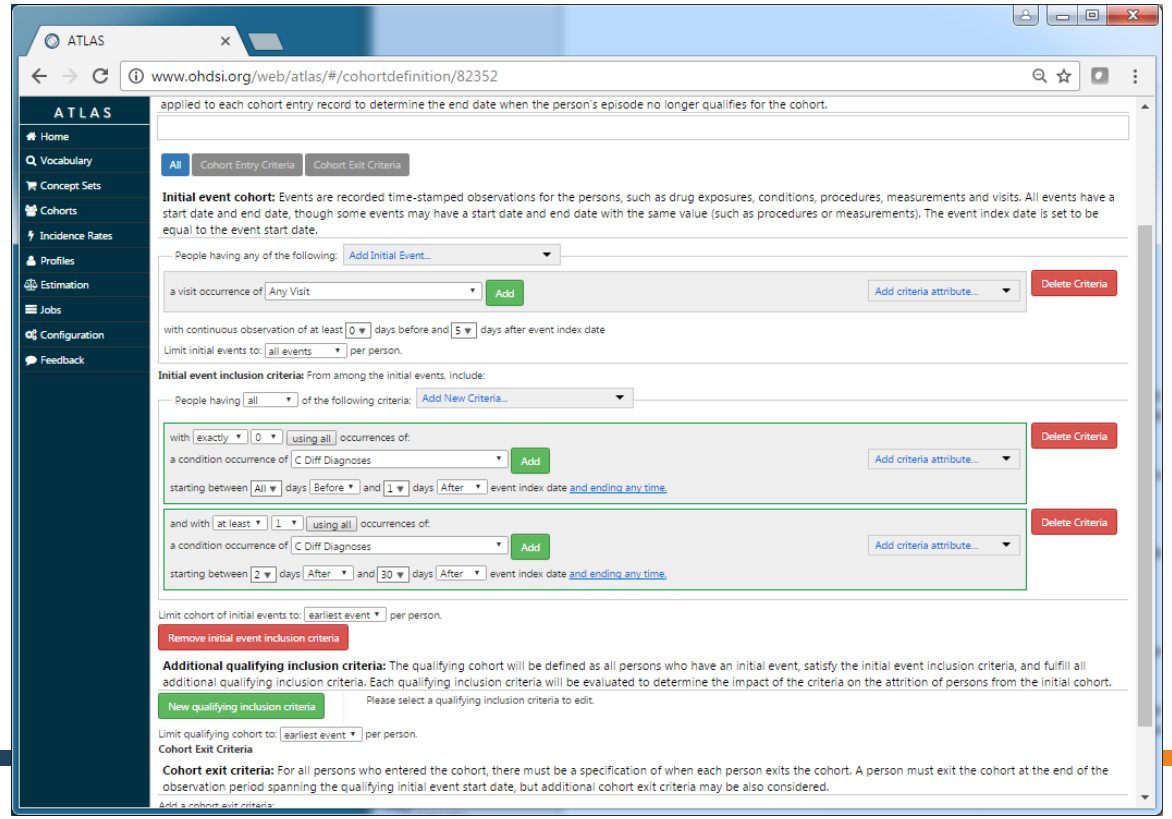
ATLAS: Ontology support

- What terms do I need to create a cohort
- Tied to the database: what terms are used
  - Especially important for someone *else's* database



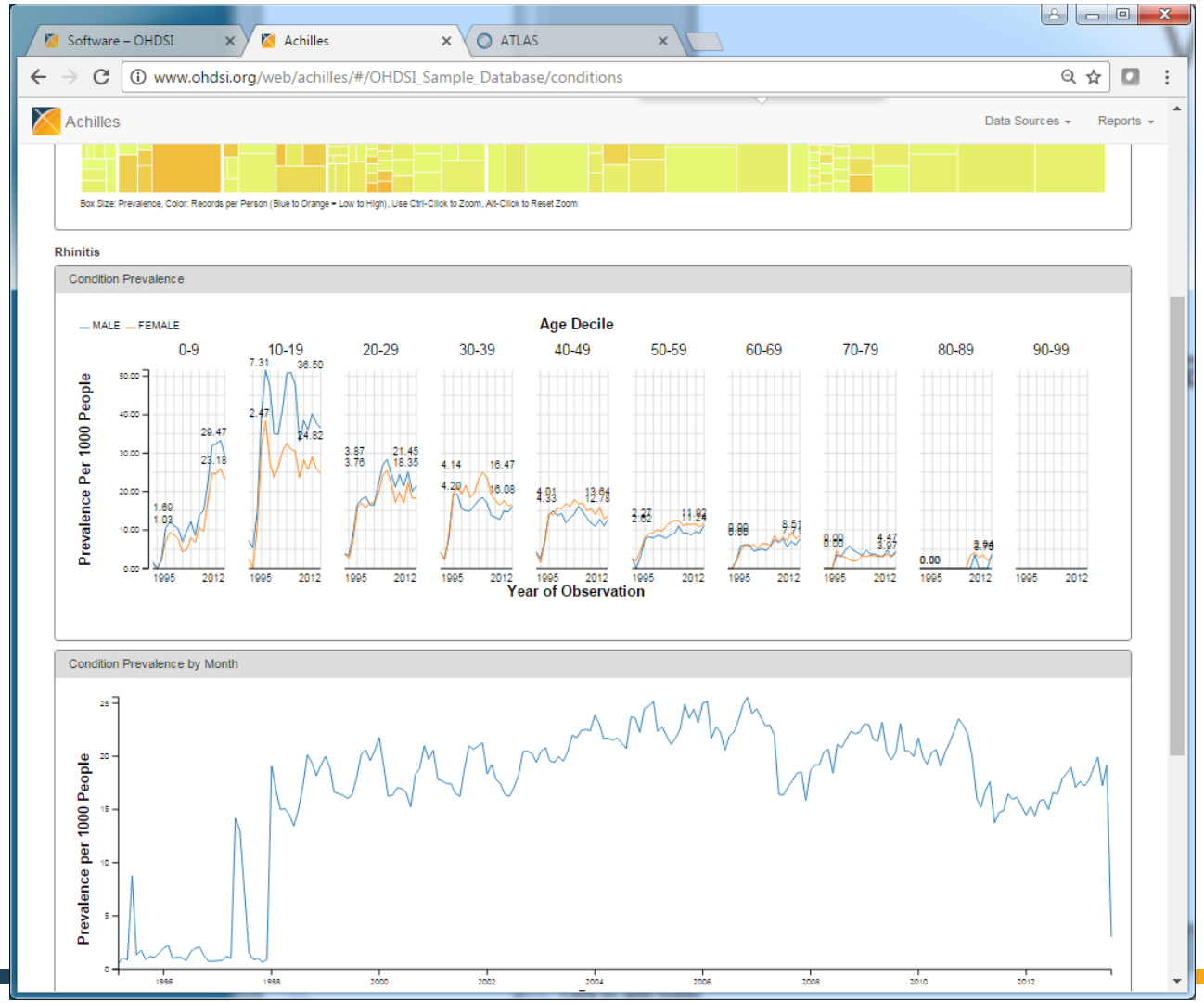
ATLAS: Cohort building

- Optimized for observational research
  - Time series: who and *when* (vs classification)
    - Observation period, event timing
  - Assume a complex definition – Linearized AND-OR group



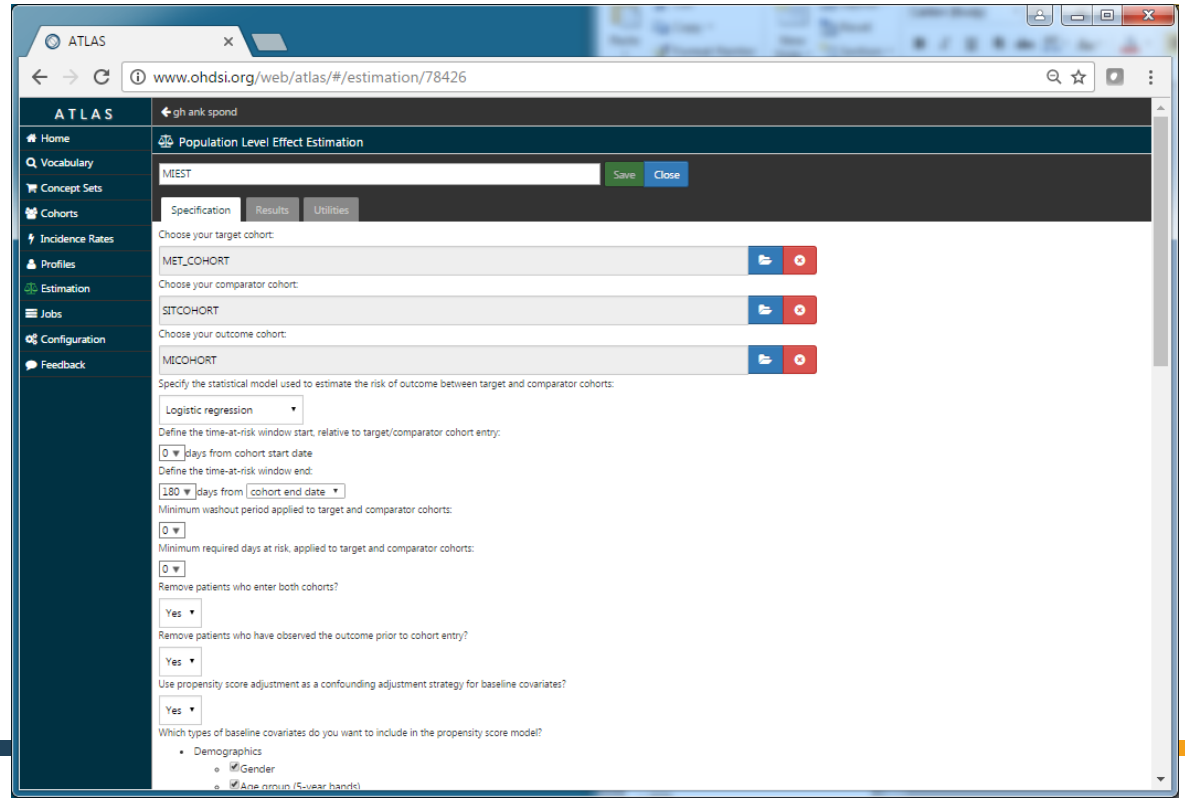
ATLAS: Visualization

- Tables
- Graphs



ATLAS: Analysis (observational)

- Approach: log regression, Poisson regression, survival
- Confounder: regularized-regression propensity score
- Residual confounding: calibration
- Diagnostics



OHDSI in Action

OHDSI “LEGEND” Hypertension Study
Filling in the evidence gaps

Clinical Practice Guideline: Executive Summary

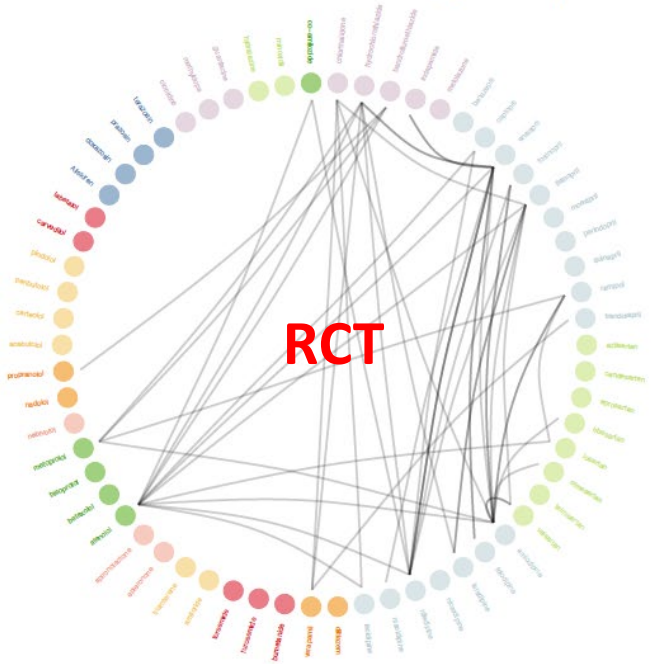
2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the Prevention, Detection, Evaluation, and Management of High Blood Pressure in Adults: Executiv

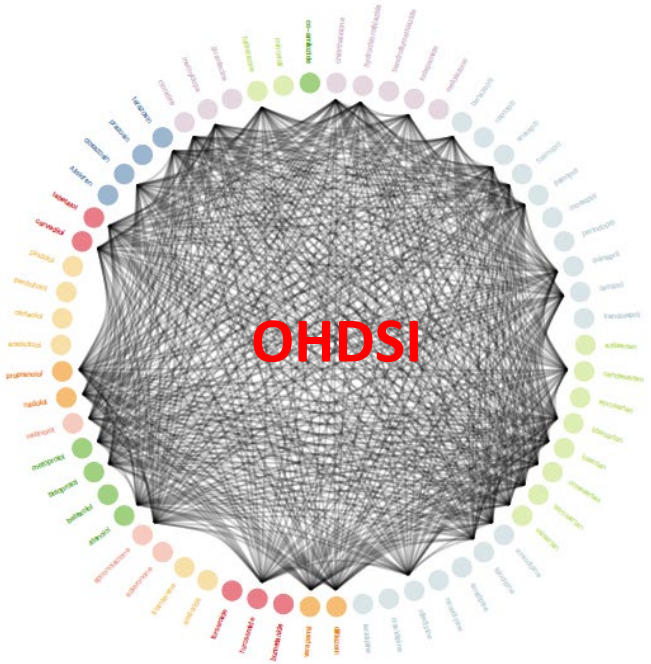
A Report of the American College of Association Task Force on Clini

WRITING COMMITTEE
Paul K. Whelton, MB, MD, MSc, FAHA, Chair; Ro
Wilbert S. Aronow, MD, FACC, FAHA*; Donald E. Casey, Jr, M
Cheryl Dennison Himmelfarb, RN, ANP, PhD, FAHA§; Sor
Samuel Gidding, MD, FAHA¶; Kenneth A. Jamerson
Eric J. MacLaughlin, PharmD**; Paul Muntner, PhD, FAHA†;
Sidney C. Smith, Jr, MD, MACC, FAHA††; Crystal C. S
Sandra J. Taler, MD, FAHA§§; Randal J. Thomas, MD, MS, FACC
Jeff D. Williamson, MD, MHS¶¶; Jackson T.

ACC/AHA TASK FORCE
Glenn N. Levine, MD, FACC, FAHA, Chair; Patrick T.
Jonathan L. Halperin, MD, FACC, FAHA, Immediate Past Cha
Joshua A. Beckman, MD, MS, FAHA; Kim K. Birtcher, MS, PharmD
Ralph G. Brindis, MD, MPH, MACC***; Joaquin E. Cigarroa
Anita Deswal, MD, MPH, FACC, FAHA; Lee A. Fleisher, M
Samuel Gidding, MD, FAHA***; Zachary D. Goldberger, MD, MS
John Ikonomidis, MD, PhD, FAHA; José A. Joglar, MD, F
Susan J. Pressler, PhD, RN, FAHA***; Barbara Riegel, PhD,

Head-to-head HTN drug comparisons

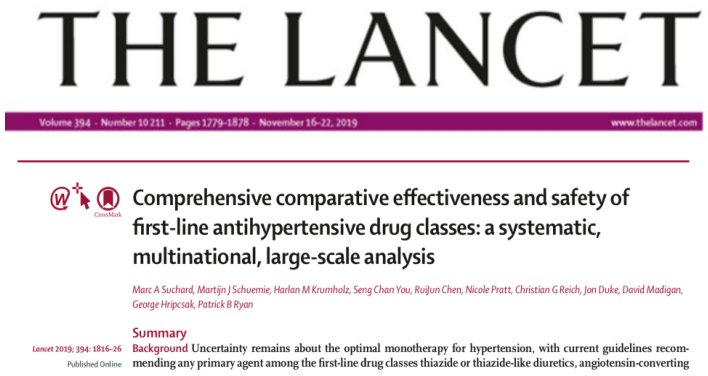




- Trials: 40
- $N = 102 - [1148] - 33K$

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

Hypertension, cardiac research



THE LANCET

Volume 394 · Number 10211 · Pages 1779-1878 · November 16-22, 2019 www.thelancet.com

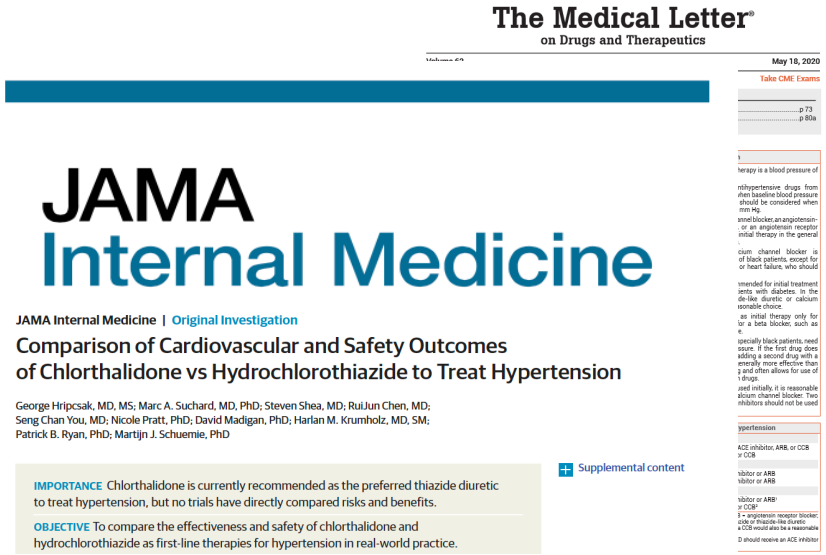
Comprehensive comparative effectiveness and safety of first-line antihypertensive drug classes: a systematic, multinational, large-scale analysis

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

Summary

Lancet 2019; 394: 1816-26 Published Online

Background Uncertainty remains about the optimal monotherapy for hypertension, with current guidelines recommending any primary agent among the first-line drug classes thiazide or thiazide-like diuretics, angiotensin-converting



The Medical Letter® on Drugs and Therapeutics

May 18, 2020

Take CME Exams

JAMA Internal Medicine

JAMA Internal Medicine | Original Investigation

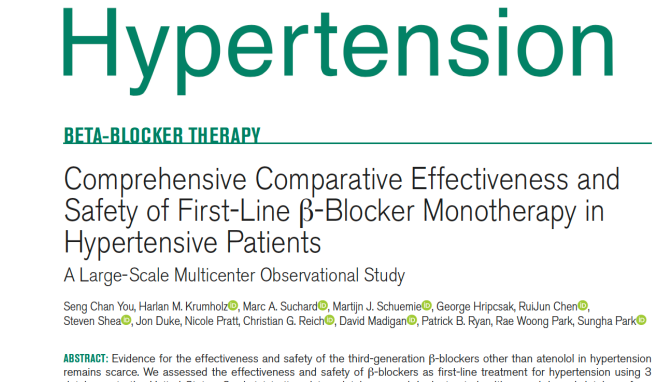
Comparison of Cardiovascular and Safety Outcomes of Chlorthalidone vs Hydrochlorothiazide to Treat Hypertension

George Hripcsak, MD, MS; Marc A. Suchard, MD, PhD; Steven Shea, MD; RuiJun Chen, MD; Seng Chan You, MD; Nicole Pratt, PhD; David Madigan, PhD; Harlan M. Krumholz, MD, SM; Patrick B. Ryan, PhD; Martijn J. Schuemie, PhD

Supplemental content

IMPORTANCE Chlorthalidone is currently recommended as the preferred thiazide diuretic to treat hypertension, but no trials have directly compared risks and benefits.

OBJECTIVE To compare the effectiveness and safety of chlorthalidone and hydrochlorothiazide as first-line therapies for hypertension in real-world practice.



Hypertension

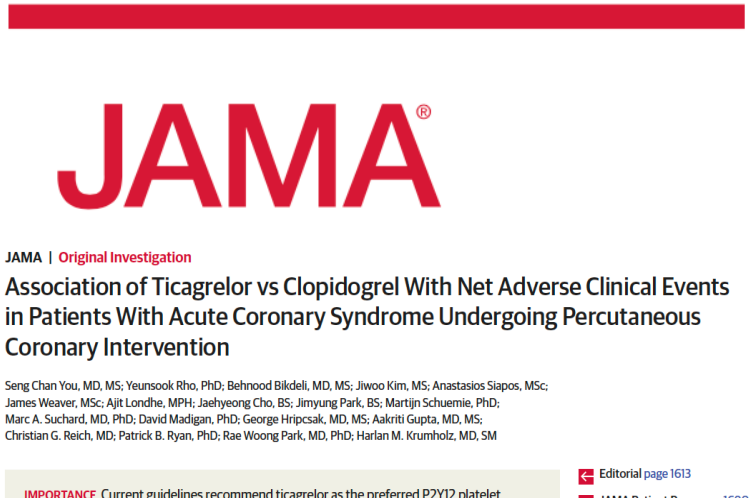
BETA-BLOCKER THERAPY

Comprehensive Comparative Effectiveness and Safety of First-Line β-Blocker Monotherapy in Hypertensive Patients

A Large-Scale Multicenter Observational Study

Seng Chan You, Harlan M. Krumholz, Marc A. Suchard, Martijn J. Schuemie, George Hripcsak, RuiJun Chen, Steven Shea, Jon Duke, Nicole Pratt, Christian G. Reich, David Madigan, Patrick B. Ryan, Rae Woong Park, Sungha Park

ABSTRACT: Evidence for the effectiveness and safety of the third-generation β-blockers other than atenolol in hypertension remains scarce. We assessed the effectiveness and safety of β-blockers as first-line treatment for hypertension using 3



JAMA®

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

Editorial page 1613

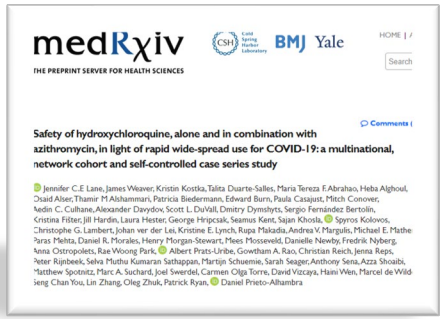
IMPORTANCE Current guidelines recommend ticagrelor as the preferred P2Y12 platelet

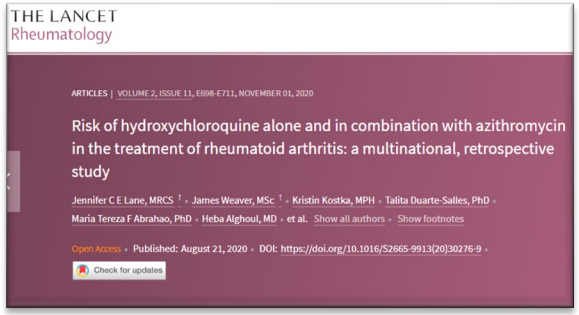
COVID-19 Research

- 4.5 million COVID-19 patient records
  - Academic centers (CDW)
  - EHR and claims data aggregators
  - Government databases
- 41 studies carried out in past year

Safety of hydroxychloroquine

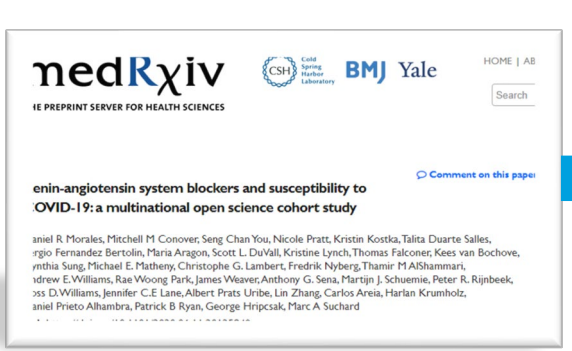
- Evidence was needed around the use of hydroxychloroquine (HCQ) alone and in combination with azithromycin (AZ). We examined the use of these drugs in rheumatoid arthritis (RA) patients.
- Findings:
  - In history use in RA population, HCQ alone is generally safe but in combination with AZ it shows a doubling of risk of 30-day cardiovascular mortality.

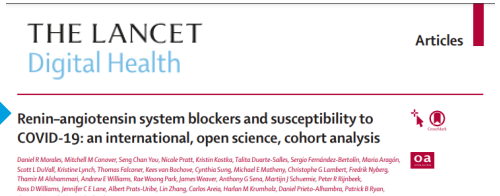




ACE Inhibitors and susceptibility to COVID-19

- Patients with cardiovascular diseases and hypertension treated with angiotensin converting enzyme inhibitors (ACEs) angiotensin-II receptor blockers (ARBs) may influence susceptibility to COVID-19 and worsen its severity.

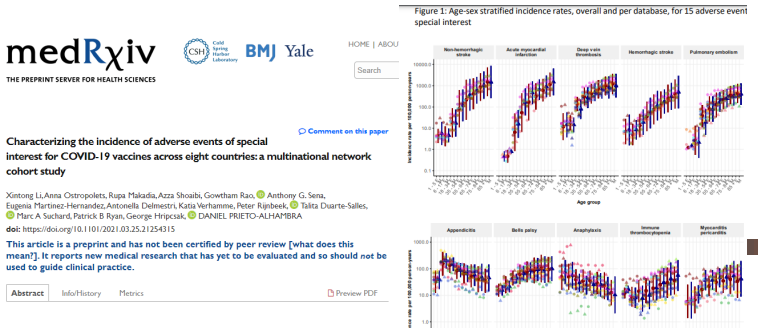


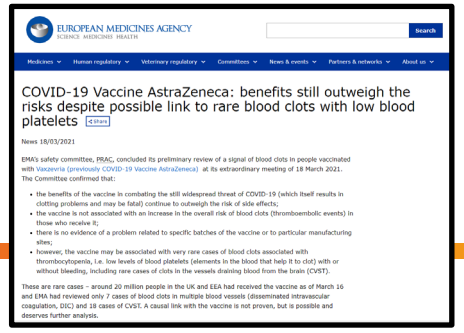


As stated by Watson et al.in relation to one of the published studies, lack of transparency and uncertainties about research standards applied raise doubts about published results. Morales et al. supported the reproducibility of their study by publishing the study protocol in the EU PAS Register ahead of time, providing a start-to-finish executable code, facilitating the sharing and exploration of the complete result set with an interactive web application and asking clinicians and epidemiologists to perform a blinded evaluation of propensity score diagnostics for the treatment comparisons.

COVID-19 Vaccine Safety Methods Research

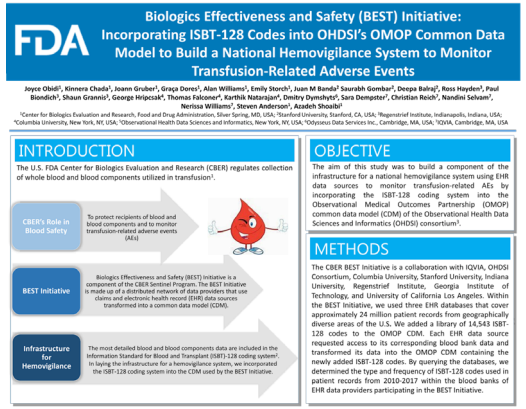
- AstraZeneca vaccine
  - March 11-15, 2021 – 13 European countries suspend use for fears of blood clots
    - Denmark, Norway, Iceland, Bulgaria, Ireland, Netherlands, Spain, Germany, Italy, France, …
  - As of March 16, 2021 – of 20 million persons vaccinated in Europe several deaths
    - 469 thromboembolic events reported after vaccination
    - 7 cases disseminated intravascular coagulation (DIC)
    - 18 cases cerebral venous sinus thrombosis (CVST)
  - March 18, 2021 – EMA determines benefits outweigh the risks
    - Thromboembolic events "lower than that **expected in the general population**"
    - DIC and CVST above baseline but very rare
    - "The number of reported events **exceeds those expected**, and causality although not confirmed, cannot therefore be excluded. However, given the rarity of the events, and the **difficulty of establishing baseline incidence** since COVID-19 itself is resulting in hospitalisations with thromboembolic complications, the strength of any association is uncertain."
- Partnered with FDA Center for Biologics Evaluation and Research (CBER)
  - Vaccine safety methods research, network and local studies



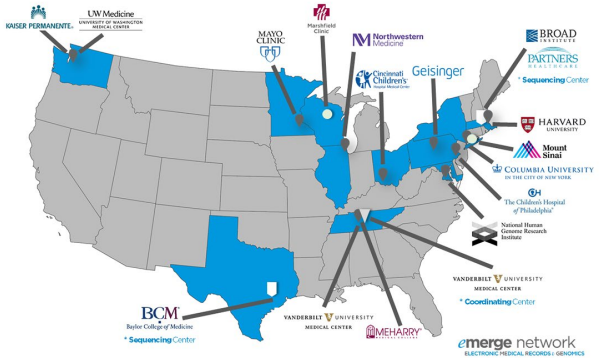


OHDSI OMOP Common Data Model

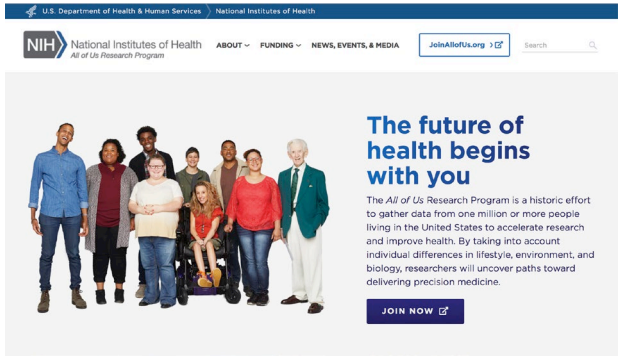
FDA CBER BEST Program



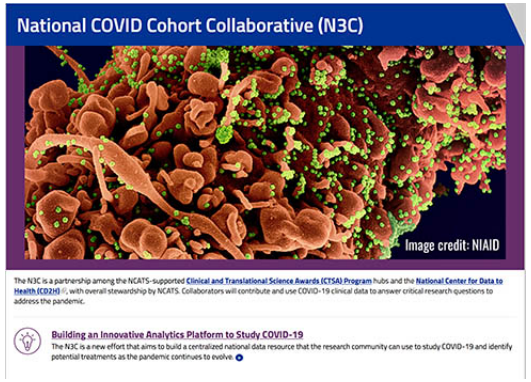
eMERGE Network



All of Us Research Program



N3C National COVID Cohort



Collaborations: PCORnet, S4S, …

HL7 – OHDSI Partnership

- Announced March 1, 2021
  - FHIR and OMOP
  - “The organizations will align their standards to capture data in a clearly defined way into a single common data model. This will allow clinicians as well as researchers to pull data from multiple sources and compile it in the same structure without degradation of the information.”
- Early in creation of joint working groups and scope
  - Starting with existing mapping work
    - Georgia Tech, EHDEN, MIRACUM, Leiden University, Denmark CSS
- Range
  - Mapping, shared knowledge engineering, common standard
- Welcome the feedback

Themes

Research as “secondary use”

- Research is not an afterthought
  - It drives health care and saves millions of lives
- Jonas Salk invented the polio vaccine
  - No one remembers his billing records
- Without research we would be billing for leeches
- Perhaps (after patient care) research should become paramount and billing should make due

Mission

- You can store data or you can generate evidence, but you cannot do both

Coupling

- Therefore, need a tightly coupled enterprise, evidence generation & standards development
  - OHDSI governance structure
  - Need analytic standards as well as data
  - Standard must be connected to community
  - Open source

It’s not magic

- Someone has to pay the data-quality price
  - In All of Us RP, found conversion is best close to data generation
  - FHIR dump of all data to a distance central warehouse will be tough

US-specific standards

- US is 5% of the population
  - Continued focus on US standards hurts US citizens
- Modern causal inference is data hungry
  - Cannot do most research on just the US population

Join the journey

http://ohdsi.org