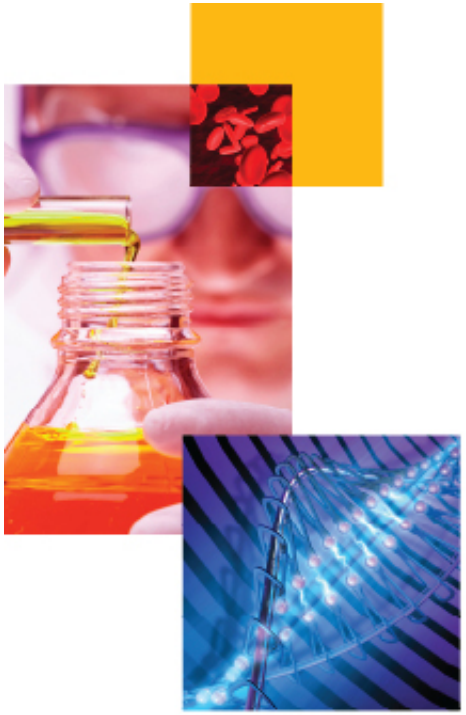


OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

FOUNDATION FOR THE National Institutes of Health

OMOP Methods Development and Evaluation

November 12, 2009

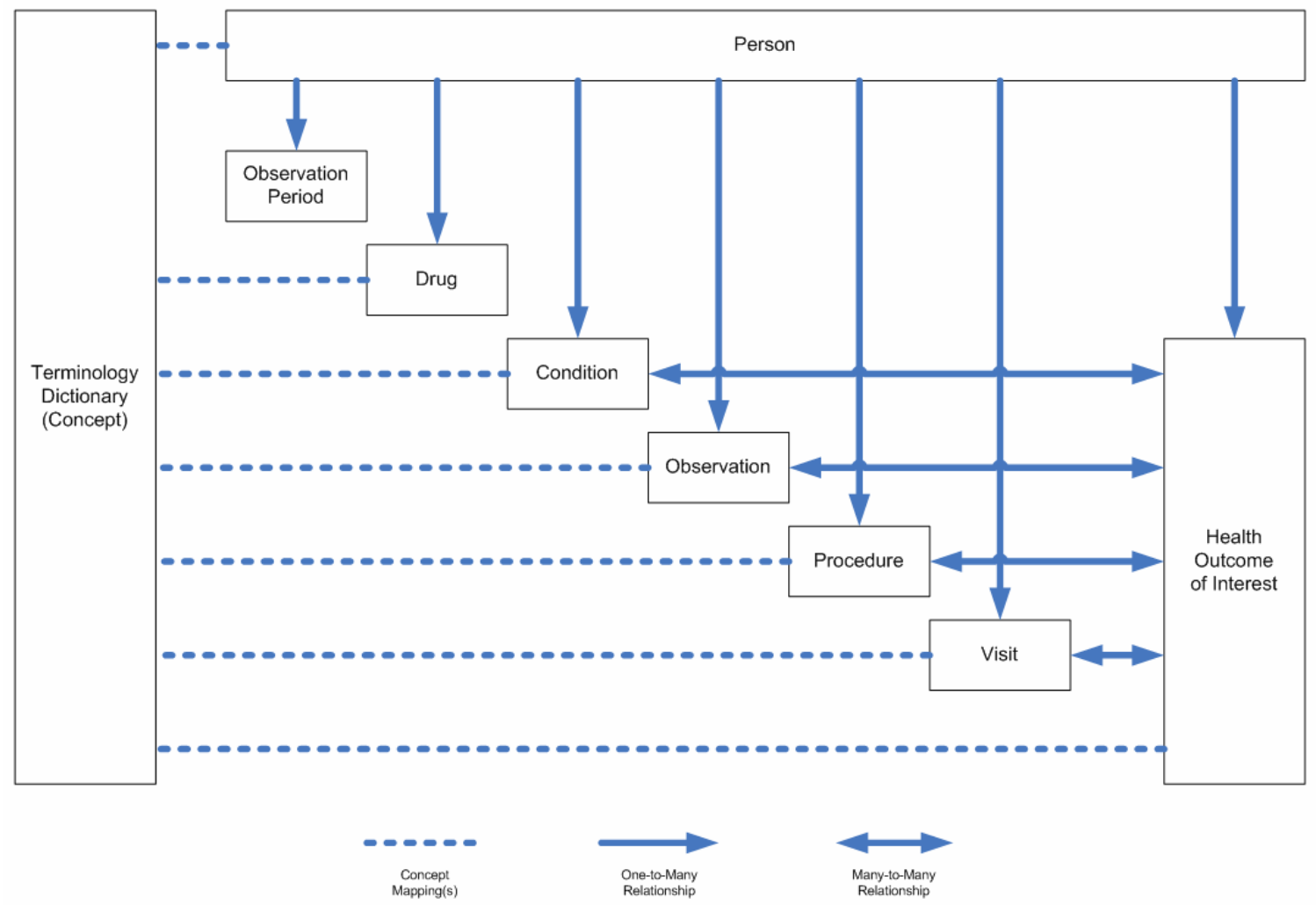
Agenda

Methods development	David Madigan, PhD
Simulating observational data	Steph Reisinger
OMOP evaluation strategy	Patrick Ryan

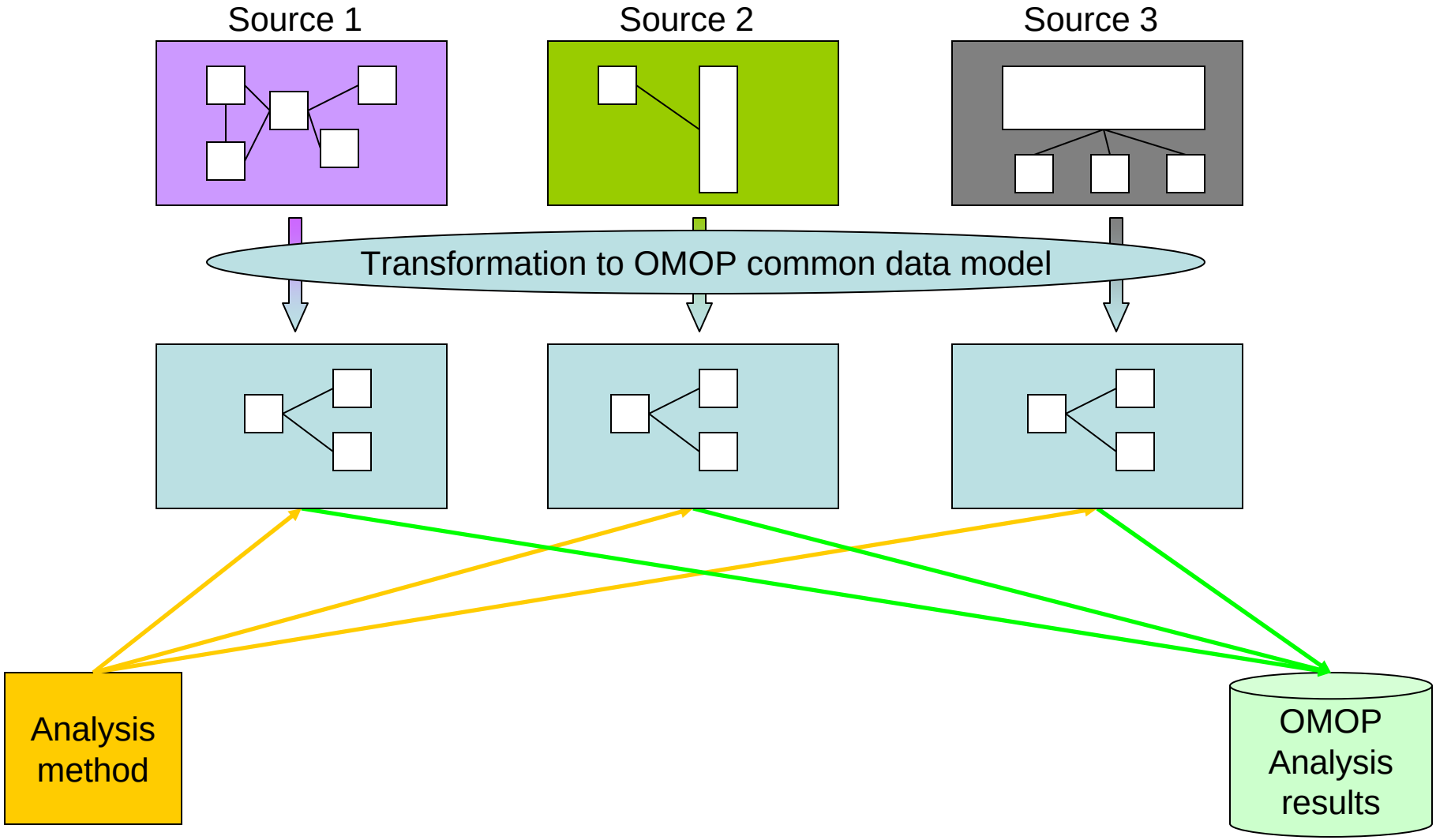
Methods development for active surveillance

David Madigan, PhD
Columbia University
OMOP Methods Lead

Conceptual Schematic of OMOP Common Data Model



Role of common data model in OMOP Analysis process



Opportunities for observational data in active surveillance

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP
FOUNDATION FOR THE National Institutes of Health

- Natural history summary of populations of interest
  - Exposed population (e.g. patients taking antibiotics)
  - Cases (e.g. patients with acute liver injury)
  - Exposed cases (e.g. patients taking antibiotics with acute liver injury)
- Case detection
- Drug-outcome associations

Characterizing Drug-Outcome Associations

Screening

Fishing expedition

Identification of non-specified conditions

Hypothesis strengthening

Prioritization

Monitoring of Health Outcomes of Interest

Data dredging

Signal substantiation

Signal detection

Exploratory analysis

Data mining

Active surveillance

Risk estimation

Signal amplification

Drug safety monitoring

Hypothesis generating

Confirmation

Evaluation

Full epi study

Signal validation

Hypothesis testing

Causal assessment

Fundamental task: *Estimate* the strength of the drug-outcome relationship

Utility and scope of methods within context of active surveillance

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP
FOUNDATION FOR THE National Institutes of Health

- Initial goal: screen to identify and prioritize drug-condition pairs which may require further evaluation

Non-specified vs Health Outcome of Interest

- Ultimate objective: elicit a valid estimate of a temporal relationship between a drug and an outcome

- No conceptual reason why any method cannot be applicable throughout the continuum of association estimation

- Practical tradeoffs: Methodological sophistication vs. scalable execution across large databases
  1. Method designed for specific exposure and outcome
  2. Method generalized for any drug and any outcome
  3. Method scalable to be applied concurrently to multiple drugs and multiple outcomes

OMOP's methods landscape

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP | FOUNDATION FOR THE National Institutes of Health

Disproportionality analysis

- Proportional reporting ratio
- Multi-item Gamma Poisson Shrinker
- Bayesian confidence propagation neural network
- Temporal pattern discovery

Other novel approaches? OMOP Cup

Exposure-based approaches

- Cohort Screening
- Incident user designs
- High dimensional propensity scoring
- Local control

Case-based approaches

- Case-control surveillance
- Case-crossover
- Self-controlled case series

Sequential methods

- Maximized sequential probability ratio test
- Conditional sequential sampling procedure

- Bayesian logistic regression
- Statistical relational learning

OMOP Methods Library at: http://omop.fnih.org/MethodsLibrary

Methodological considerations common across multiple approaches

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP
FOUNDATION FOR THE National Institutes of Health

- Exposure definition
  - Incident vs. prevalent exposure
  - Source of data capture
- Outcome definition
  - Incident vs. prevalent events
  - Diagnosis codes vs. rules across data elements
- Defining temporal relationship
  - Time from exposure start
  - Time after exposure end
- Comparator selection
- Inclusion/exclusion criteria
  - Baseline history
  - Follow-up time
- Covariate selection and adjustment
  - Matching
  - Stratification
  - Clustering
  - Multivariate modeling
- Output metric/statistic
  - Test threshold vs. effect estimate
  - Relative vs. attributable risk
  - Measure of uncertainty

Simulated Observational Data

Steph Reisinger
ProSanos Corporation

Context for Simulated Data

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP | FOUNDATION FOR THE National Institutes of Health

"The OMOP partnership will conduct a two-year initiative to research methods that are feasible and useful to analyze existing healthcare databases to identify and evaluate safety and benefit issues of drugs already on the market"

source: omop.fnih.org

- **Method Evaluation Requirements:**
  - **Characterize Individual Method Performance**
  - **Compare Performance Among Methods**

Objective Measures for:

- Sensitivity
- Specificity
- Positive and Negative Predictive Value

Limitations of Observational Data for Supporting Methods Evaluation

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP | FOUNDATION FOR THE National Institutes of Health

Observational Data	Simulated Data
Data is “noisy” (confounding)	Model both adverse drug reactions and confounding
Data capture process provides further distortion	Simulate data capture process
Limited gold standards for objective measurement	Known characteristics provide “truth” for measurement
Access limited & expensive	Data freely & widely available
Disparate data formats / coding schemes	Use of Common Data Model mitigates format issues

Simulated data, with known properties and characteristics, can facilitate systematic evaluation & comparison among methods providing objective gold standard

Observational Medical Dataset Simulator

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP
FOUNDATION FOR THE National Institutes of Health

- **OSIM**
  - Developed to address need for objective Method evaluation criteria
  - Open source software application (R)
  - File of hypothetical persons with fictitious drug exposures and condition occurrences
  - Input parameters mimic characteristics of real observational data
  - OMOP "Common Data Model" format
  - **Support OMOP Methods Development & Evaluation and *Methods Cup* competition**

OSIM Conceptual Design

- **Sufficiently represent complexity of drug event associations in real observational data**
- **Include features important for method evaluation**

Drugs & Conditions

- **Fictitious**
- **Broad Range of Characteristics**

Health Experiences

- **Model of Population Characteristics**

Healthcare Encounters

- **Simulate Recording of Observational Data**

Drug and Condition Key Characteristics

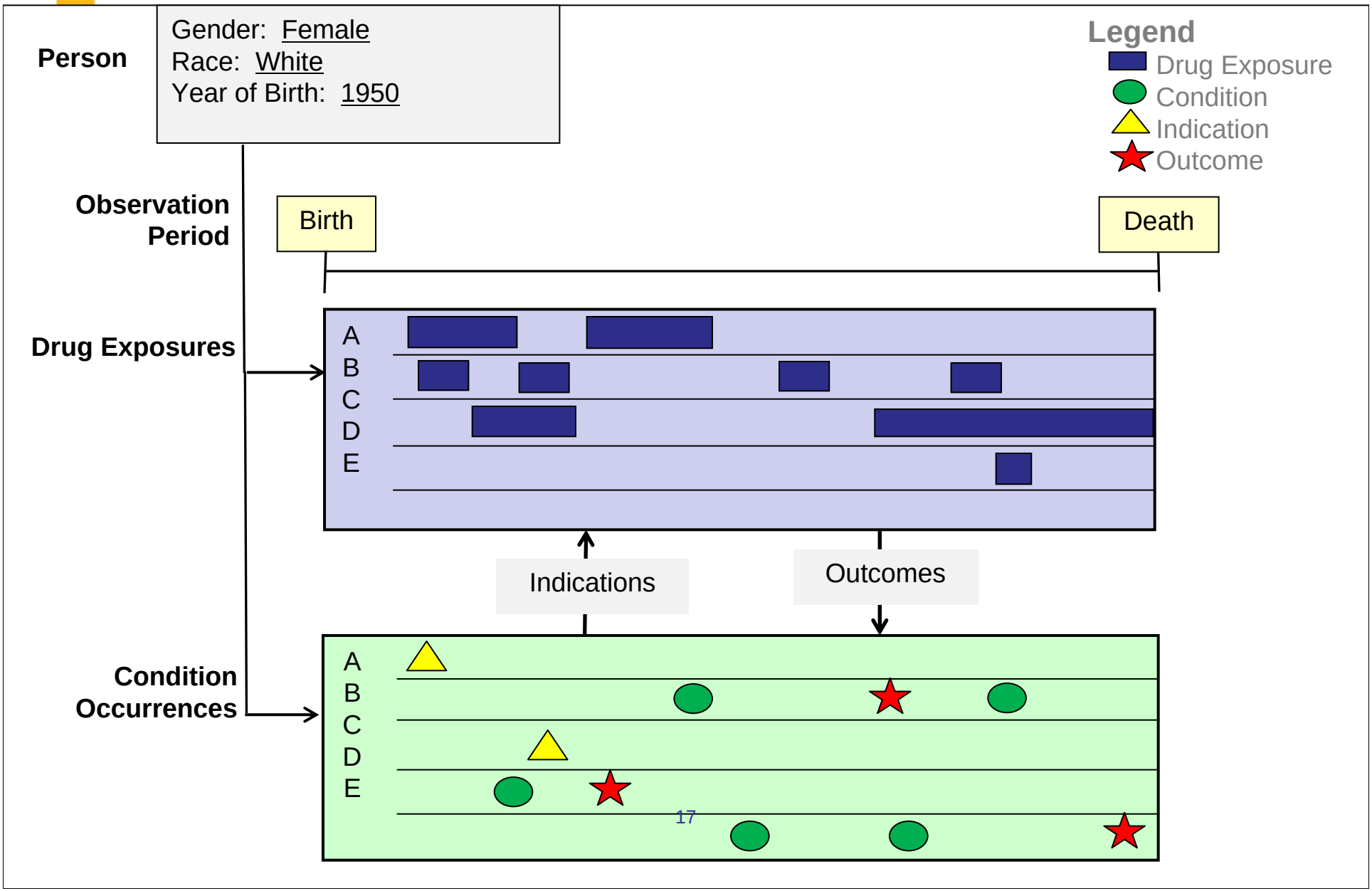
OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP | FOUNDATION FOR THE National Institutes of Health

Drugs & Conditions

- **Fictitious**
- **Broad range of characteristics**

Category	Key Characteristics
Drugs	• Background Prevalence • Difference due to Age, Gender, Race, other Confounders • Exposure frequency and length
Conditions	• Background Prevalence • Difference due to Age, Gender, Race, other Confounders
Indications	• Condition → Drug Relationships
Drug Outcomes	• Drug → Condition Relationships • Time to Onset

Simulating Health Experiences



Health Experiences Key Characteristics

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP | FOUNDATION FOR THE National Institutes of Health

Health Experiences

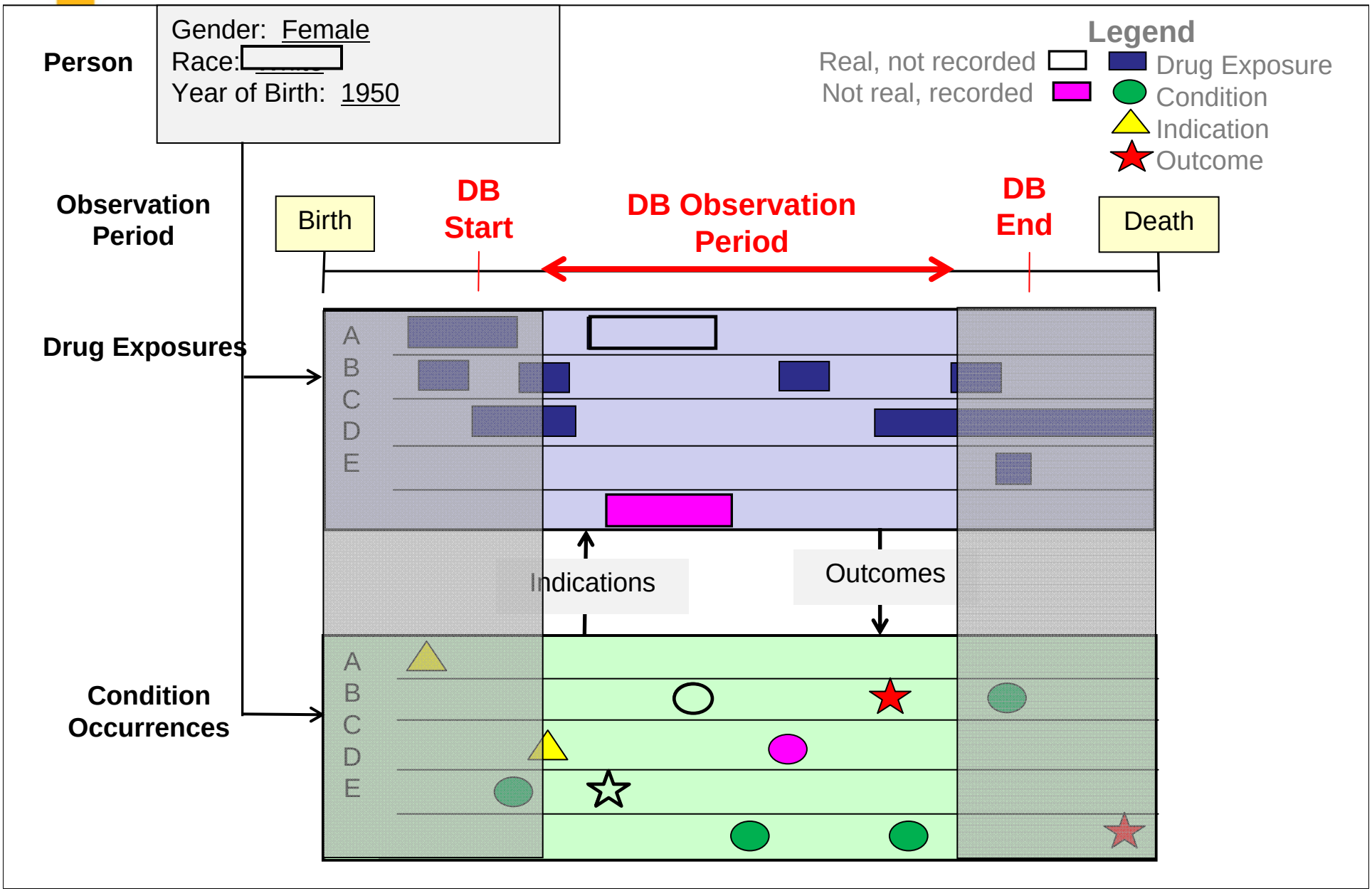
- **Model population characteristics**

Category	Key Characteristics
Person	• Age • Patient Demographics • Other Confounding Factors

Condition Occurrences	• Assignment of Condition Occurrences
	  • Background rates
	  • Increased / decreased risk due to Person Confounders
	  • Drug-related outcomes

Drug Exposures	• Assignment of Drug Exposures
	  • Background rates
	  • Increased / decreased risk due to Person Confounders
	  • Indication related Exposures
	  • Length & frequency

Simulating Healthcare Encounters



Healthcare Encounter Key Characteristics

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP | FOUNDATION FOR THE National Institutes of Health

Healthcare Encounters

- **Simulate Recording of Observational Data**

Category	Key Characteristics

Person	• Recording Person Data
	  • Observation Period
	  • Patient Demographics Missing / Recorded

Condition Occurrences	• Recording Condition Occurrences
	  • Data Capture Sensitivity
	  • Data Capture Specificity

Drug Exposures	• Recording Drug Exposures
	  • Data Capture Sensitivity
	  • Data Capture Specificity
	  • Indication Recorded Proportion and Placement

Addressing Confounding

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP
FOUNDATION FOR THE National Institutes of Health

Factors independently related to exposure and outcome which could affect assessment of drug-related outcomes

- **Confounding in OSIM**
  - Model nominal background prevalence for all drugs and conditions
  - Randomly introduce confounding into select drug / condition relationships
    - Select sub-populations may have different likelihoods of drug exposure or condition occurrence
  - Indications introduced in excess of background
  - Drug-related outcomes introduced in excess of background

OSIM Functional Requirements

Drugs & Conditions

Health Experiences

Healthcare Encounters

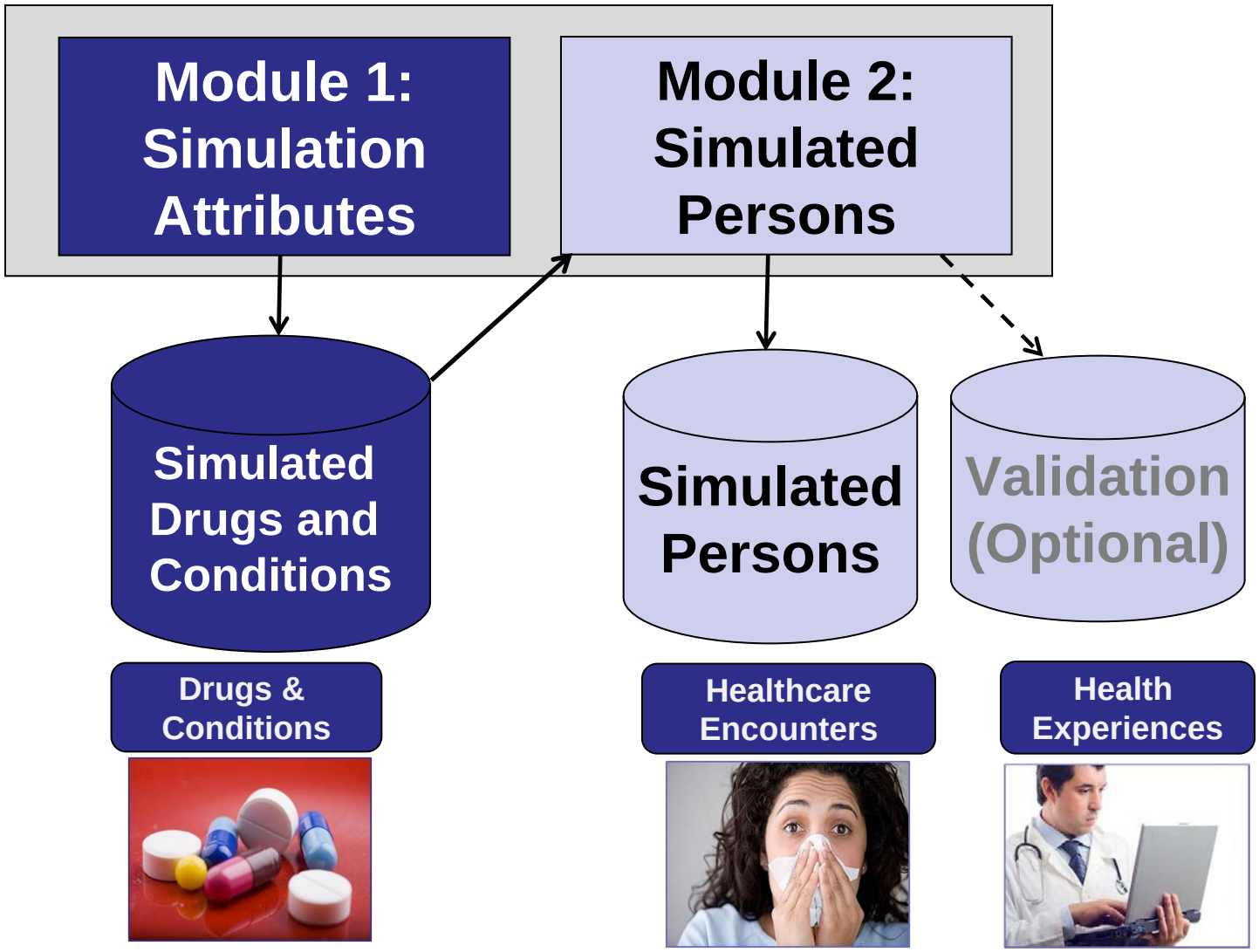
- **User can control simulated data characteristics using input parameters**
- **Simulated Drug and Condition output kept separate from Simulated Persons output**
  - File containing “truth” and ability to blind an analysis
- **Scale to 100,000,000 Simulated Persons**
- **Conform to OMOP Common Data Model**

Input Parameters

- **User can control the characteristics of the simulated data through 30+ input probability distributions**
- **Default distributions developed based on characteristics of real obs. data**

Person Age Probability Distribution
Person Age Recorded Probability Distribution
Condition Age Attributable Risk Probability Distribution
Drug Age Attributable Risk Probability Distribution
Person Gender Probability Distribution
Person Gender Recorded Probability Distribution
Condition Gender Attributable Risk Probability Distribution
Drug Gender Attributable Risk Probability Distribution
Person Race Probability Distribution
Person Race Recorded Probability Distribution
Condition Race Attributable Risk Probability Distribution
Drug Race Attributable Risk Probability Distribution
Person Confounder Probability Distribution
Condition Confounder Attributable Risk Probability Distribution
Drug Confounder Attributable Risk Probability Distribution
Person Age at Death Probability Distribution
Observation Period Probability Distribution
Condition Baseline Prevalence Probability Distribution
Condition Data Capture Sensitivity Probability Distribution
Condition Data Capture Specificity Probability Distribution
Condition Occurrence per Person Probability Distribution
Drug Baseline Prevalence Probability Distribution
Drug Data Capture Sensitivity Probability Distribution
Drug Data Capture Specificity Probability Distribution
Drug Exposure Length Probability Distribution
Drug Number of Exposures Probability Distribution
Drug Prior Indication Probability Distribution
Drugs per Indication Probability Distribution
Indication Outcome Attributable Risk Probability Distribution
Drug Outcome Attributable Risk
Drug Outcome Time to Onset

Separate Output



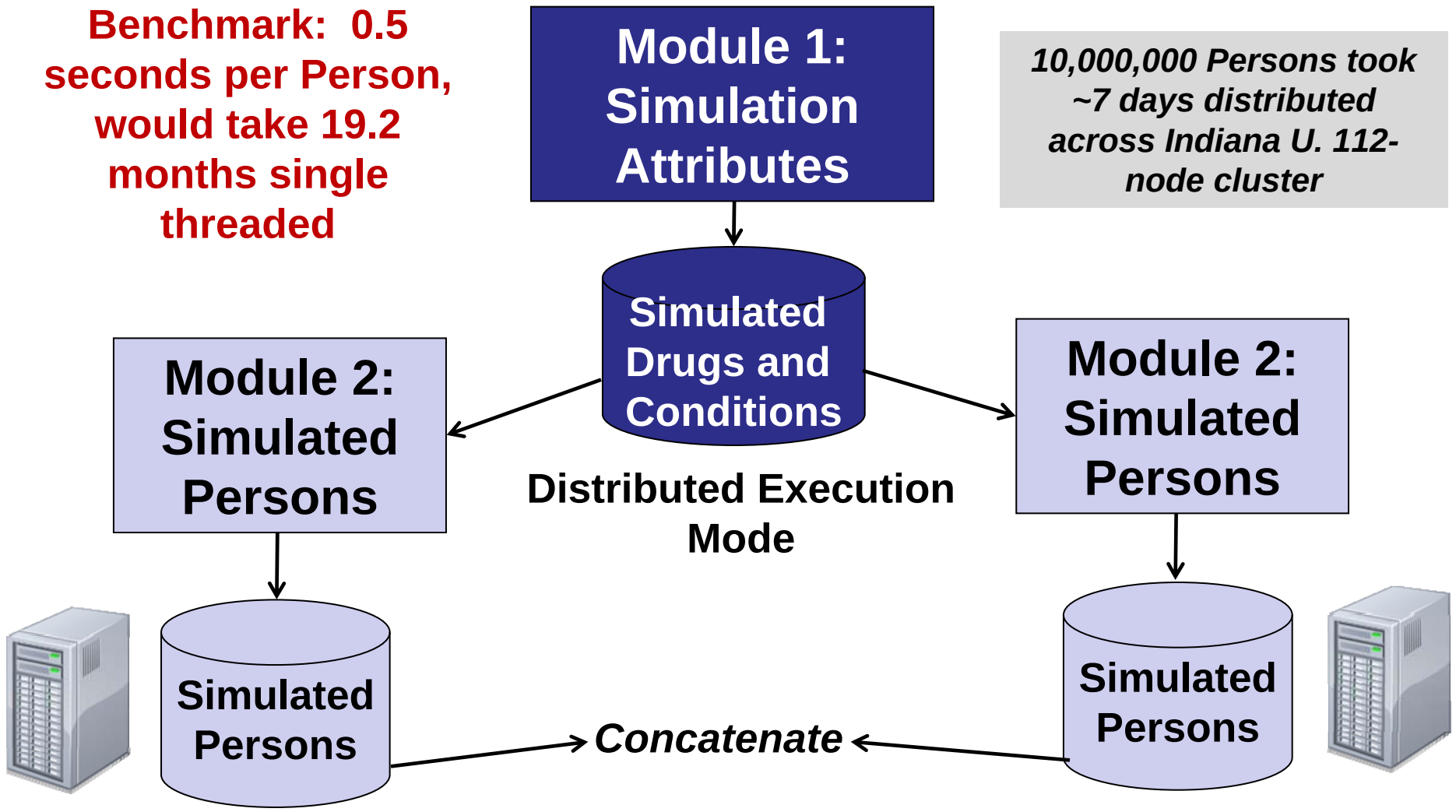
All health experiences, and why they were or weren't recorded in the Simulated Persons Data

100,000,000 Person Simulation

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP | FOUNDATION FOR THE National Institutes of Health

Theoretical Benchmark: 0.5 seconds per Person, would take 19.2 months single threaded

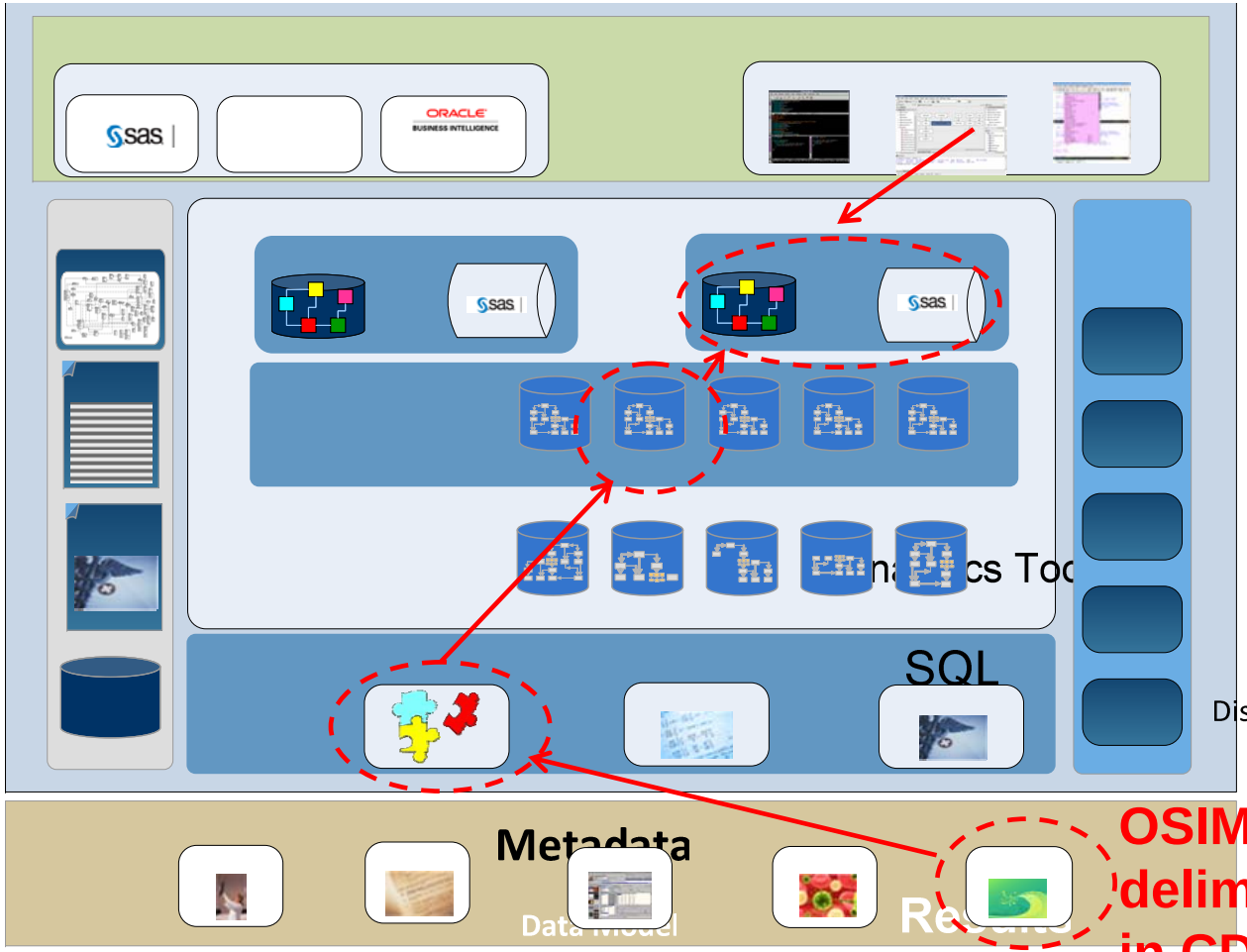
10,000,000 Persons took ~7 days distributed across Indiana U. 112-node cluster



OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP | FOUNDATION FOR THE National Institutes of Health

OMOP Lab

Load into Oracle Database



Access by Methods Library for Method Evaluation

Analytics Serv

Discoverer

OMOP Data

OSIM produces delimited text files in CDM format

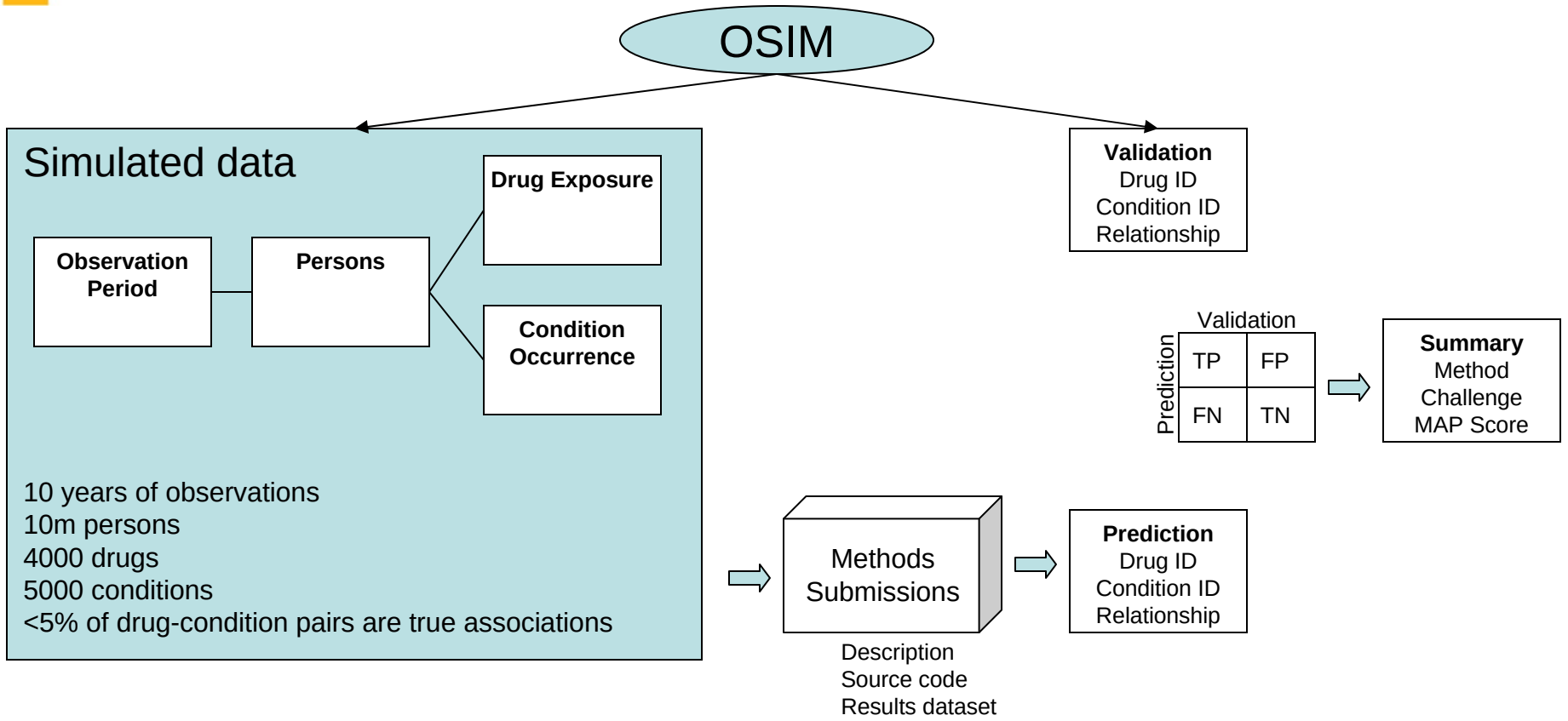
Summary: OSIM Strengths & Limitations for Methods Research

Strengths	Limitations
Enables objective methods evaluation using data that is well characterized	Oversimplification of clinical characteristics and relationships
Provides objective baseline for methods comparison (sensitivity, specificity, positive and negative predictive value)	Not ideal for hypothesis Strengthening Methods where study design and covariate selection rely on clinical expertise
Impact Parameters can be varied and the impact studied	Does not simulate procedures, visits, or other observations
Can be used for hypotheses generation and monitoring health outcomes of interest	

OMOP Evaluation Strategy

Patrick Ryan
GlaxoSmithKline
OMOP Research Investigator

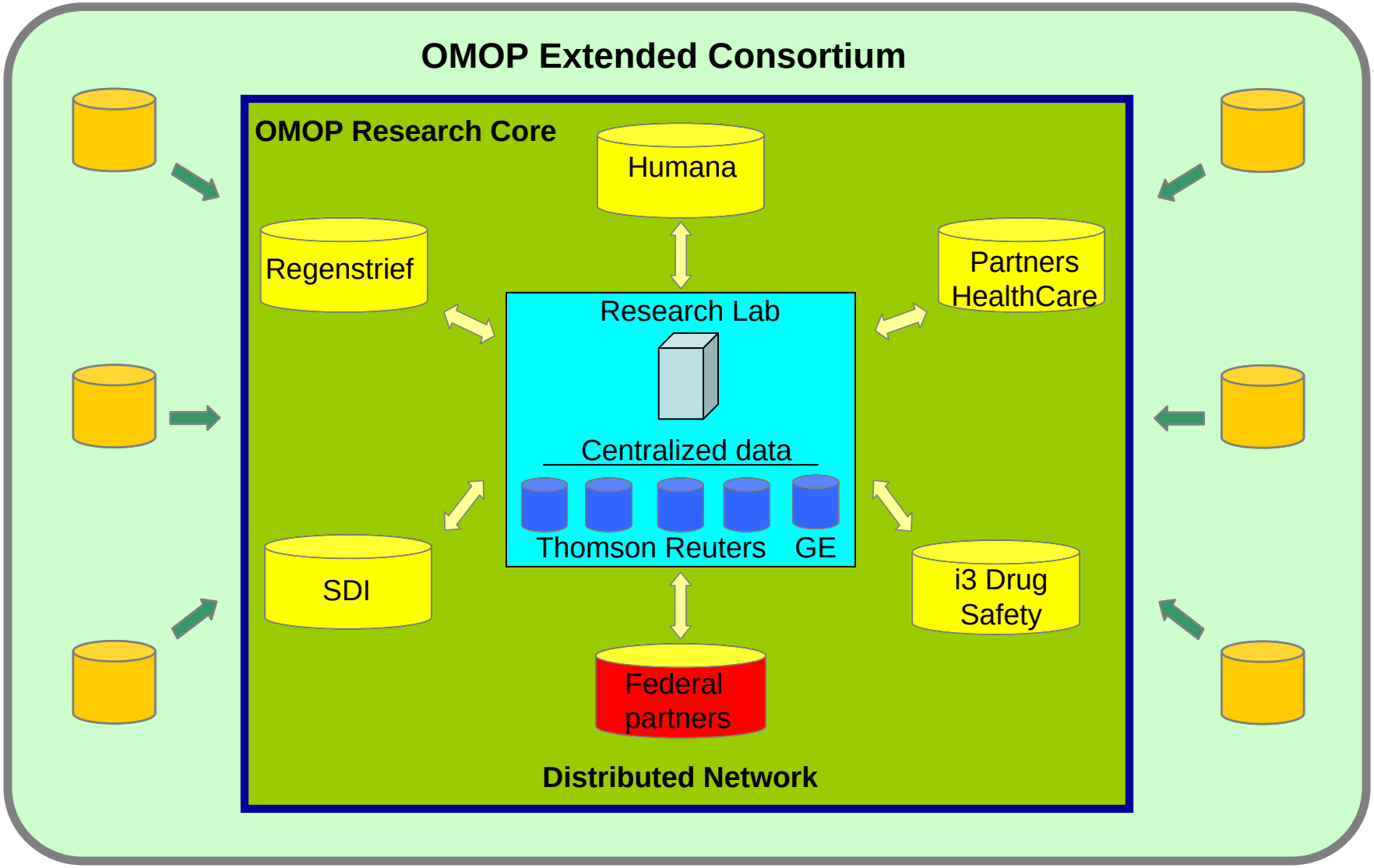
OMOP Cup: Methods Competition



- Two competitions: http://omopcup.orwik.com
  - Challenge 1: Identifying drug-condition associations within an entire observational dataset
  - Challenge 2: Identifying drug-condition associations as data accumulates over time
- Evaluation criteria: Weighted Mean Average Precision
- Winning entries will be given cash prize and methods will be further tested against OMOP data community

OMOP Data Community

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP
FOUNDATION FOR THE National Institutes of Health



Methods testing strategy: Monitoring of Health Outcomes of Interest

- Each Health Outcome of Interest has one or more operational definitions that are applied to each database to determine cases for analysis
- Each method will be implemented in the OMOP Research Lab against the central databases
- Feasible methods will be tested across the OMOP data community
- Methods performance tested in two dimensions
  - Identifying drug-condition associations within an entire observational dataset
  - Identifying drug-condition associations as data accumulates over time
- Evaluation focuses on degree to which method maximizes ‘true positives’ while minimizing ‘false positives’
- Monitoring of Health Outcomes of Interest studies for each method will a pre-defined set of relationships (true association and negative control) between 10 HOIs and 10 drugs

Drug-Health Outcomes of Interest pairs under study

Drug/class	Health Outcome of Interest
ACE inhibitors	Angioedema
ACE inhibitors	Hospitalization (including readmission and mortality)
Amphotericin B	Renal failure
Antibiotics: erythromycins, sulfonamides, and tetracyclines	Acute liver injury (symptomatic hepatitis)
Antiepileptics: carbamazepine, valproic acid, and phenytoin	Aplastic anemia
Benzodiazepines	Hip fracture
Beta blockers	Mortality after MI
Bisphosphonates: alendronate	GI ulcer hospitalizations
Tricyclic antidepressants	Myocardial infarction
Typical antipsychotics	Myocardial infarction
Warfarin	Bleeding

Methods testing strategy: Identification of non-specified conditions

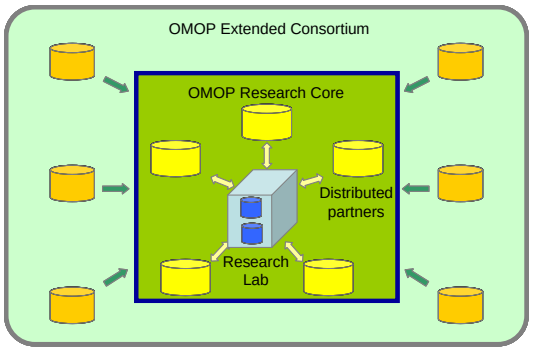
OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP
FOUNDATION FOR THE National Institutes of Health

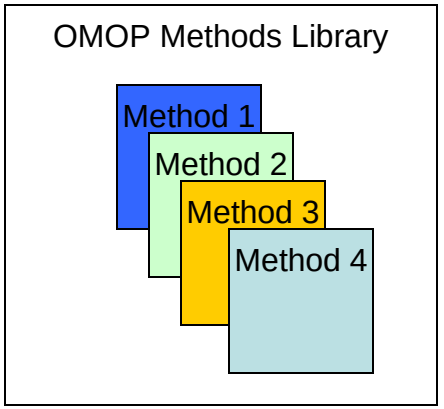
- Each method will be implemented in the OMOP Research Lab against the central databases, and tested against simulated data and across the OMOP data community
- Methods performance tested within an entire observational dataset and as data accumulates over time
- Studies across OMOP data community will explore all outcomes for 10 drugs and compare to 'labeled events'
  - 'Labeled event' extracted from structured product labels through natural language processing program developed by Regenstrief
  - 'Labeled events' characterized by where they are listed on label
    - Warning
    - Precautions
    - Adverse Reactions
    - Postmarketing Experience
  - For purposes of methodological research, 'labeled events' will be classified as 'true' associations and conditions unrelated to 'labeled events' in the the reference condition ontology will be classified as 'negative controls'. Any association identified by the method that is not a 'true' association is to be considered a 'false positive' and will not be further reviewed.

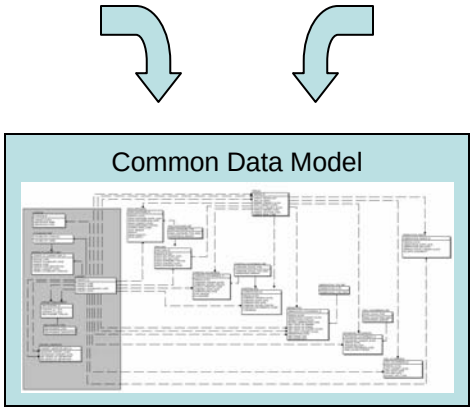
Evaluating the performance of methods: Mean Average Precision

		Original Values			Sorted Values		
Drug	**Condition**	z_i	y_i		$z_{(i)}$	$y_{(i)}$	$P^{(K)}$
D1	C1	5	1		9	1	1/1=1
	C2	0	1		8	1	2/2=1
	C3	9	1		7	0	
D2	C1	8	1		5	1	3/4=0.75
	C2	4	1		4	1	4/5=0.8
	C3	3	0		3	0	
D3	C1	1	0		2	0	
	C2	2	0		1	0	
	C3	7	0		0	1	5/9=0.55
						Total Score	(1+1+0.75+0.8+0.55)/5 =0.82

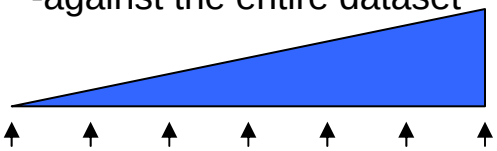
Methods experiment workflow







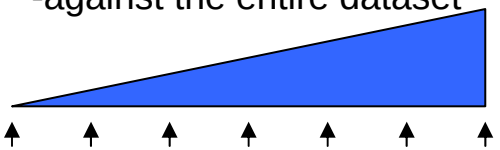
Testing in each source:
-accumulating over time
-against the entire dataset



Health Outcomes of Interest
1. Angioedema
2. Aplastic Anemia
3. Acute Liver Injury
4. Bleeding
5. GI Ulcer Hospitalization
6. Hip Fracture
7. Hospitalization
8. Myocardial Infarction
9. Mortality after MI
10. Renal Failure

Drugs
1. ACE Inhibitors
2. Amphotericin B
3. Antibiotics
4. Antiepileptics
5. Benzodiazapines
6. Beta blockers
7. Bisphosphonates
8. Tricyclic antidepressants
9. Typical antipsychotics
10. Warfarin

Testing in each source:
-accumulating over time
-against the entire dataset



Non-specified conditions

-All outcomes in condition terminology

-'Labeled events' as reference
-Warning
-Precautions
-Adverse Reactions
-Postmarketing Experience

Open discussion