

# Observational Medical Outcomes Partnership – Industry Perspective

Jonathan Morris MD  
United BioSource



# Disclaimer



The views and opinions expressed in the following PowerPoint slides are those of the individual presenter and should not be attributed to Drug Information Association, Inc. (“DIA”), its directors, officers, employees, volunteers, members, chapters, councils, Special Interest Area Communities or affiliates, or any organization with which the presenter is employed or affiliated.

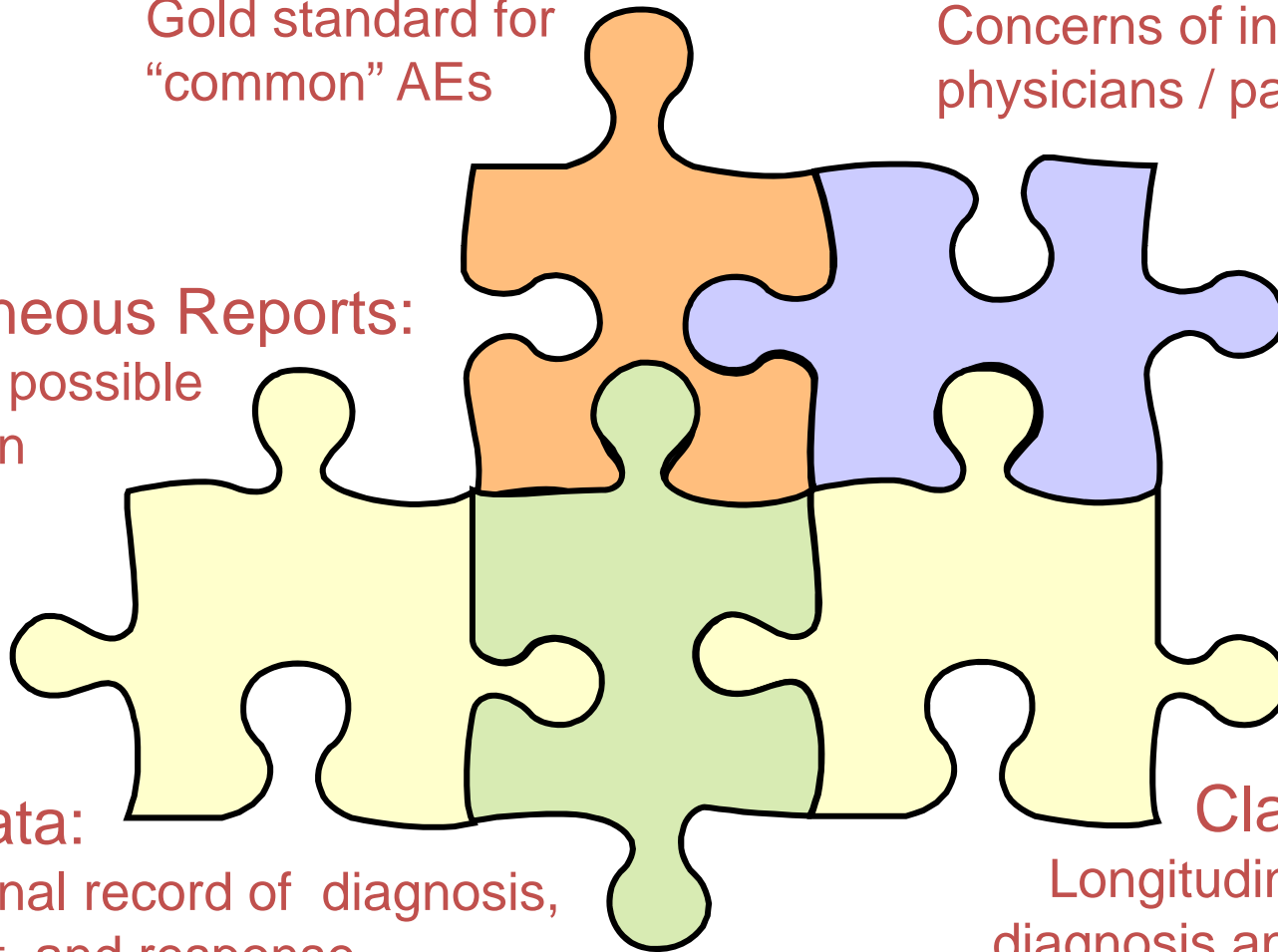
These PowerPoint slides are the intellectual property of the individual presenter and are protected under the copyright laws of the United States of America and other countries. Used by permission. All rights reserved. Drug Information Association, DIA and DIA logo are registered trademarks or trademarks of Drug Information Association Inc. All other trademarks are the property of their respective owners.

- **OMOP Organizational Overview**
- OMOP Areas of Focus
  - Data Partners and Simulated Data
    - Data Characteristics, Data Quality
    - Common Data Model
    - Ontologies
  - Cohort Selection
  - Health Outcomes of Interest
  - Analytic Methods
- **Summary**

**Clinical Trials:**  
Gold standard for  
“common” AEs

**Case Reports:**  
Concerns of individual  
physicians / patients

**Spontaneous Reports:**  
Broadest possible  
population



**EHR Data:**  
Longitudinal record of diagnosis,  
treatment, and response

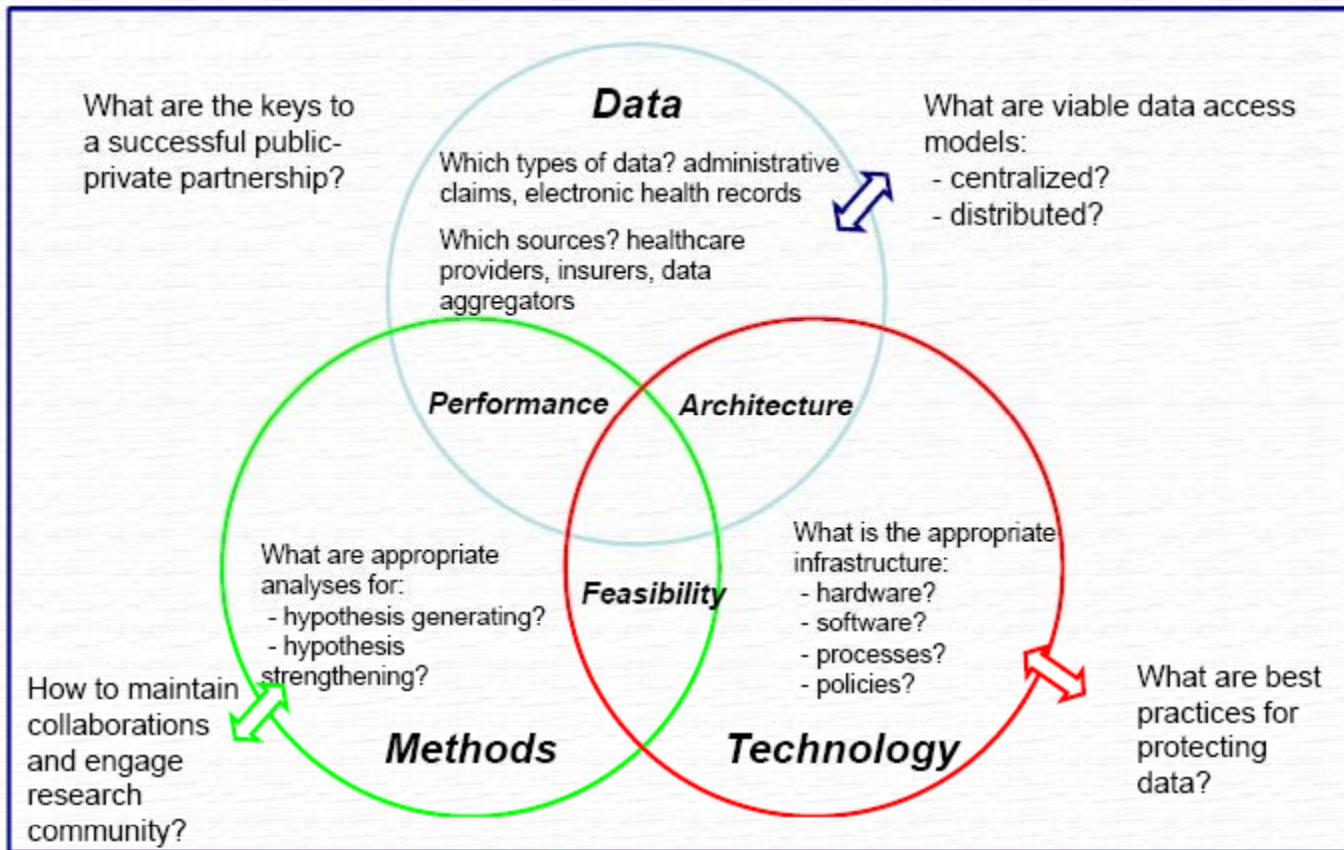
**Claims Data:**  
Longitudinal record of  
diagnosis and treatment

***A public-private partnership to serve the public health by testing whether multi-source observational data can improve our ability to assess drug safety and benefits.***

- Assess the appropriate technology and data infrastructure required for systematic monitoring of observational data
- Develop and test the feasibility and performance of the analysis methods
- Evaluate required governance structures

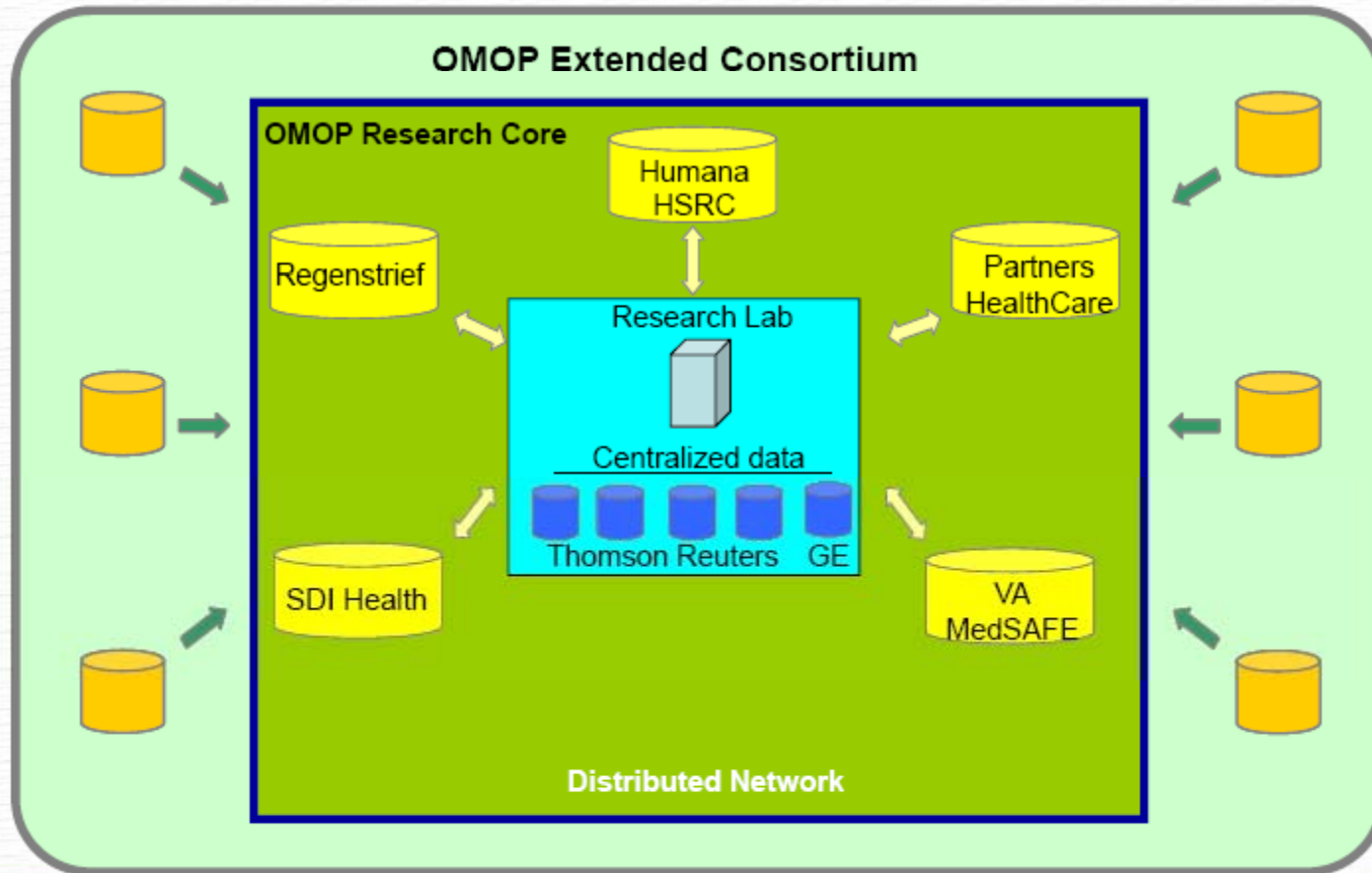
- Natural history summary of populations of interest
  - Exposed population (e.g. patients taking antibiotics)
  - Cases (e.g. patients with acute liver injury)
  - Exposed cases (e.g. patients taking antibiotics with acute liver injury)
- Case detection
- Drug-outcome associations
- Lower cost
- More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

# Outstanding Questions for Active Surveillance



- **Phase 1: FEASIBILITY OF DATA INFRASTRUCTURE (Feb – July 2009)**
  - Establish a consistent framework to use across disparate observational data sources
  - Establish OMOP Research Community
- **Phase 2: FEASIBILITY OF ANALYSES (Aug – Dec 2009)**
  - Develop and test analysis methods within the OMOP Research Lab and other data environments
  - Establish standard data characterization procedures
  - Implement health outcomes of interest definitions
  - OMOP to facilitate comparisons across databases
- **Phase 3: PERFORMANCE MEASUREMENTS (Jan – July 2010)**
  - Evaluate performance of methods and data in identifying drug safety issues
  - OMOP to facilitate comparisons across databases
- **Phase 4: UTILITY OF ANALYSES & PROCESS (July – Dec 2010)**
  - Assess the effectiveness and usefulness of how the results and comparisons contribute to decision-making

- *OMOP Organizational Overview*
- OMOP Areas of Focus
  - **Data Partners and Simulated Data**
    - Data Characteristics, Data Quality
    - Common Data Model
    - Ontologies
  - Cohort Selection
  - Health Outcomes of Interest
  - Analytic Methods
- Summary



## Centralized Database Characteristics



	CCAЕ	MDCD	MDCR	MSLR	GE
Type of DB	Claims	Claims	Claims	Claims	EHR
Data Characteristics	Commercial Claims	Medicaid	Medicare	Claims Supplemented with Lab data	Primary Care Providers
Number of Patients	46,456,798	10,779,424	45,558,58	1,229,321	11,216,208
Lab Values?	No	No	No	Yes	Yes

**All 5 databases transformed into OMOP CDM prior to methods research**



## Observational Data

Data is “noisy” (confounding)

Data capture process provides further distortion

Limited gold standards for objective measurement

Access limited & expensive

Disparate data formats / coding schemes

## Simulated Data

Model both adverse drug reactions and confounding

Simulate data capture process

Known characteristics provide “truth” for measurement

Data freely & widely available

Use of Common Data Model mitigates format issues

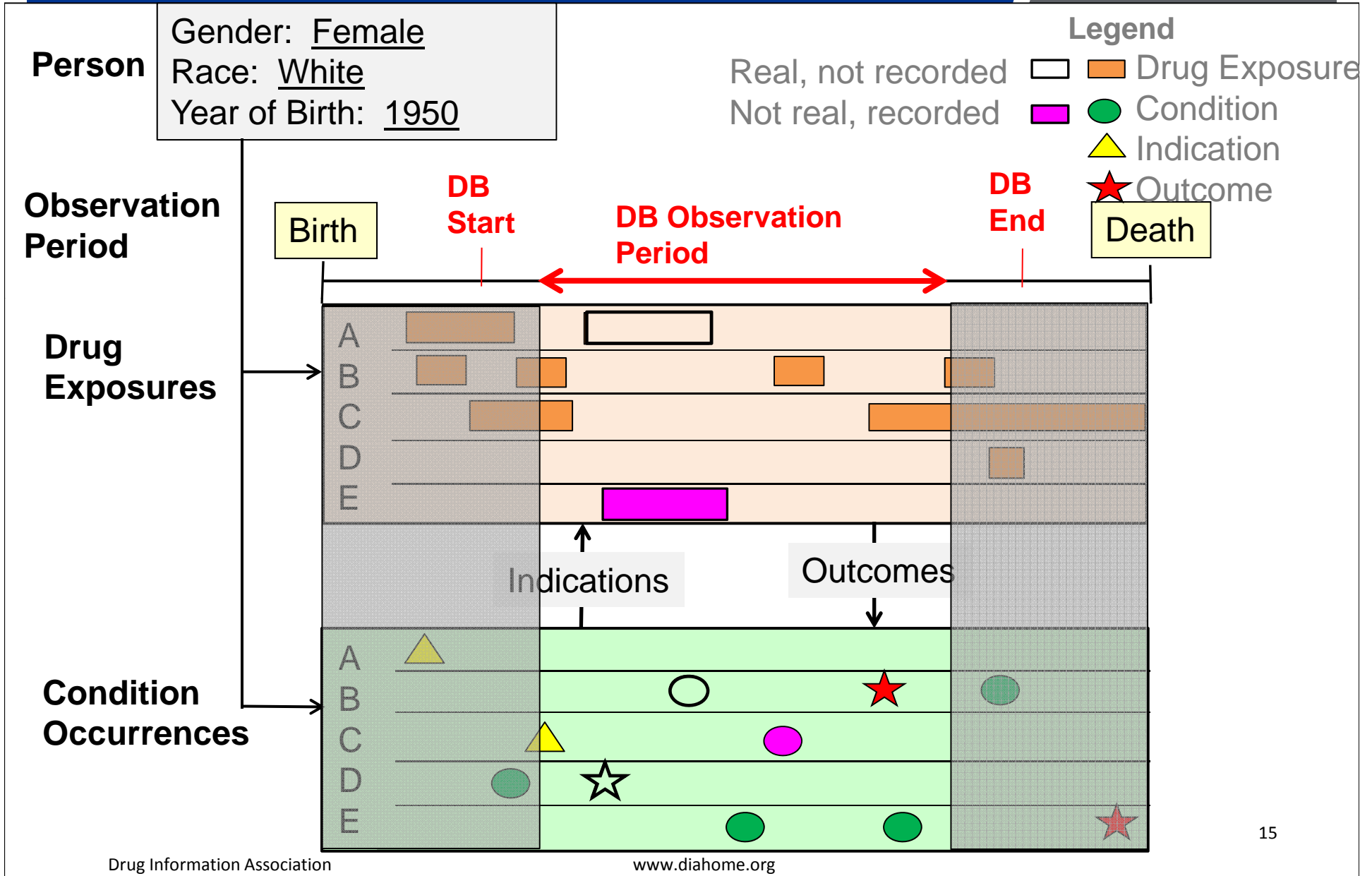
**Simulated data, with known properties and characteristics, can facilitate systematic evaluation & comparison among methods providing objective gold standard**

- OSIM I
  - First generation of observational data simulation program
  - Successfully used for method development and enabled OMOP cup
  - Modeled first-order univariate effects (population demographics, prevalence, distribution of drugs and conditions)

- OSIM II

- Alternative design to accommodate additional complexities, including relationships between / among conditions and drugs
- Generates 12 “transition probability tables” from real observational data representing the probability of identified data characteristics
  - Represent derived characteristics and confounding, e.g. # of drugs per person, correlation between conditions, correlations between conditions and drugs prescribed, co-prescribing, etc.
- Simulated data generated in CDM format based on values/probabilities found within the transition probability tables

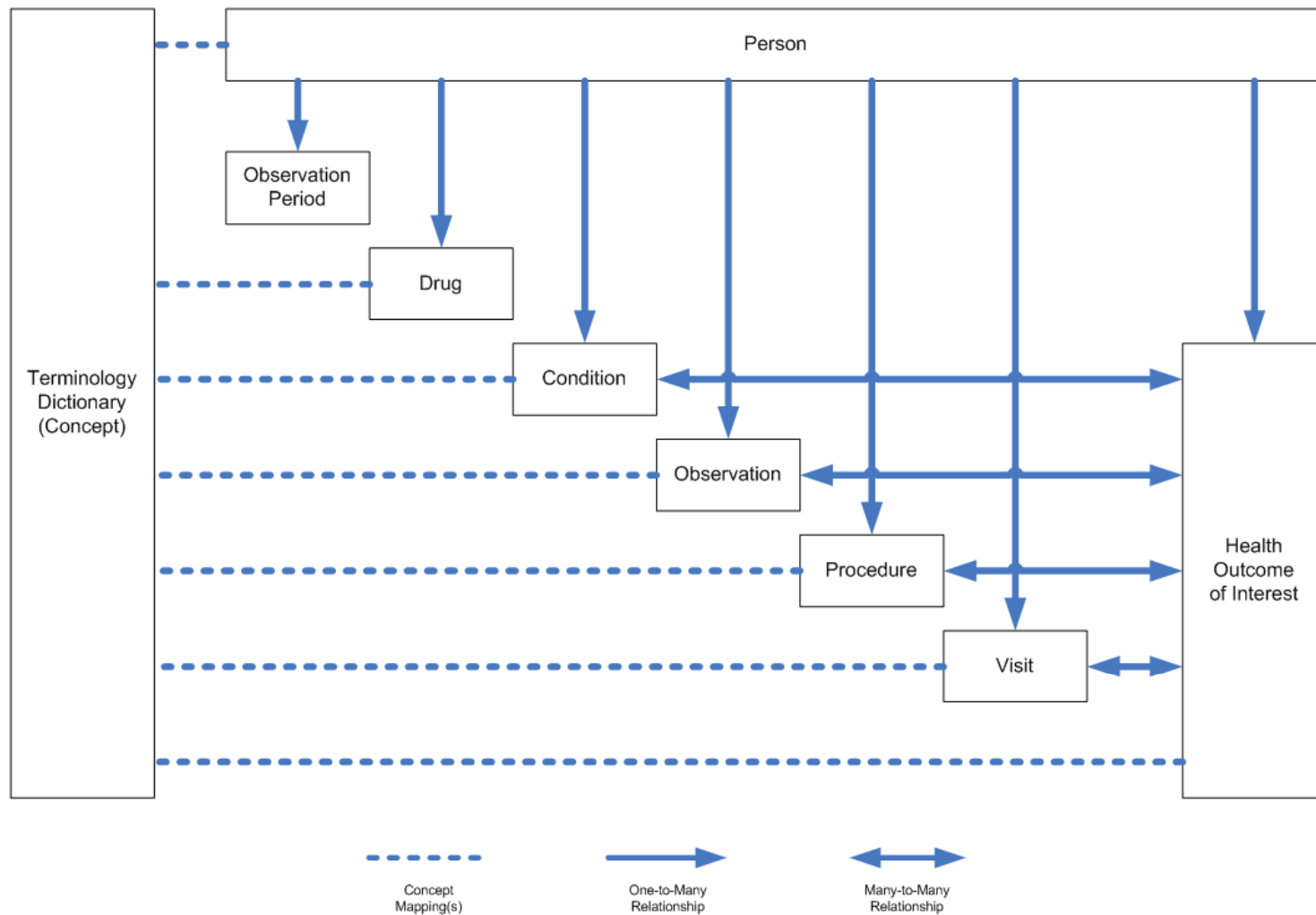
# Simulating Reality



- The establishment of the network of diverse data holding organizations yielded insights into the capabilities needed to be a successful active surveillance site
  - Minimum **technology requirements** to support analysis of large-scale databases
  - Staff requirements needed to set up and operate an active surveillance site
  - The importance of investigators at each partner site to contribute to the **understanding of the variability** seen from one data source to another
  - Good **governance** – the need to balance transparency with participation
  - Use of simulated data is a valuable addition to complement methodological research

- *OMOP Organizational Overview*
- *OMOP Areas of Focus*
  - *Data Partners and Simulated Data*
    - *Data Characteristics, Data Quality*
    - **Common Data Model**
    - *Ontologies*
  - **Cohort Selection**
  - **Health Outcomes of Interest**
  - **Analytic Methods**
- **Summary**

# Conceptual Schematic of OMOP Common Data Model



# What does observational data look like?

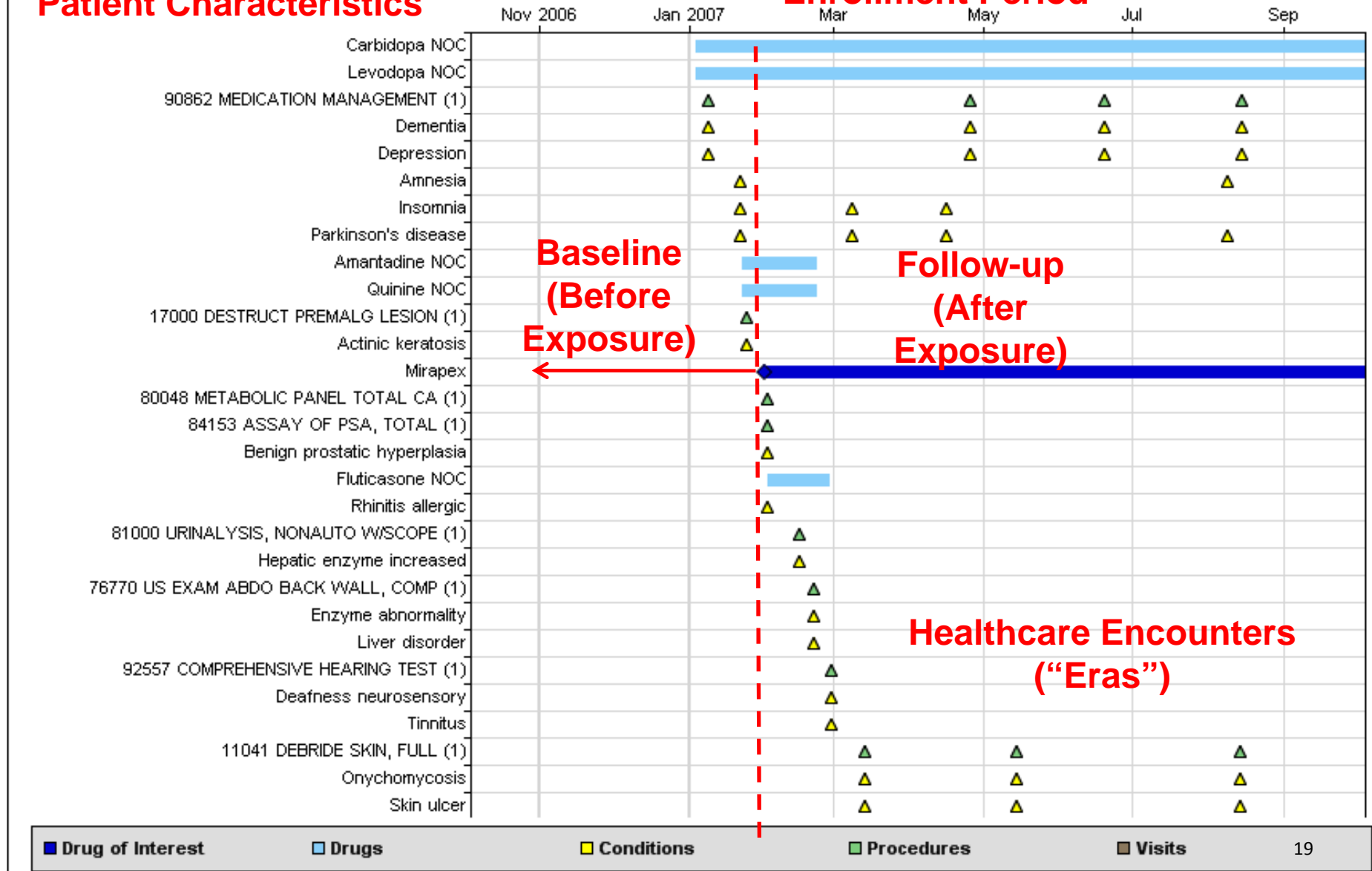


Person ID: 0261AAAAAABAPZHM

Birth Year: 1935 Age: 72 Gender: male Race: Region: W

## Patient Characteristics

## Enrollment Period



- Provides a systematic approach for summarizing observational healthcare data stored in OMOP common data model.
- Uses
  - Validation of transformation from raw data to OMOP common data model
  - Comparisons between data sources
  - Comparison of overall database to specific subpopulations of interest (such as people exposed to particular drug or people with specific condition)
  - Providing context for interpreting and analyzing findings of drug safety studies

<http://omop.fnih.org/OSCAR>

- Data Quality program that produces a summary report for each data source of warnings of implausible and suspicious data observed from the OSCAR summary.
- It identifies potential issues across all OMOP common data model tables, including potential concerns with all drug exposures and all conditions.
  - Data quality review of specific drugs (such as the ingredients that comprise the OMOP drugs of interest).
  - Data quality review of specific conditions (including population-level prevalence of the health outcomes of interest, and unexpected gender-specific rates, such as males with pregnancy, and females with prostate cancer).

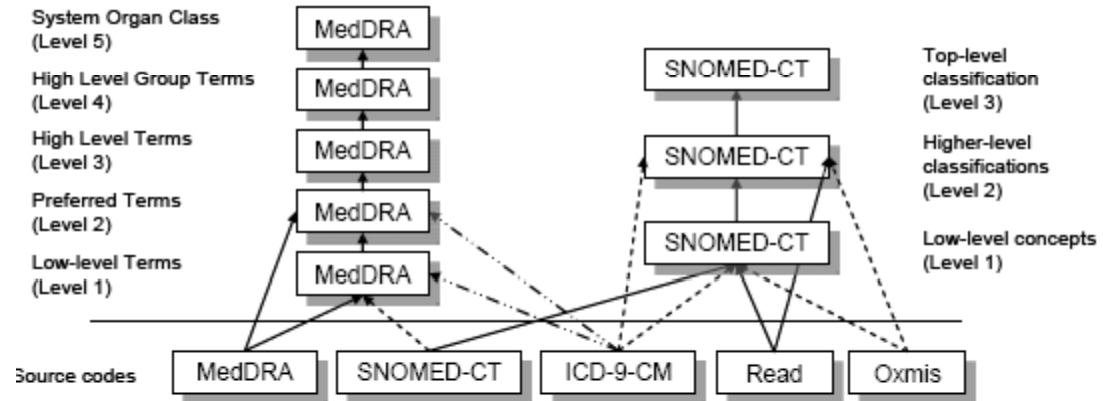
<http://omop.fnih.org/GROUCH>

- *OMOP Organizational Overview*
- *OMOP Areas of Focus*
  - *Data Partners and Simulated Data*
    - *Data Characteristics, Data Quality*
    - *Common Data Model*
    - **Ontologies**
  - Cohort Selection
  - Health Outcomes of Interest
  - Analytic Methods
- **Summary**

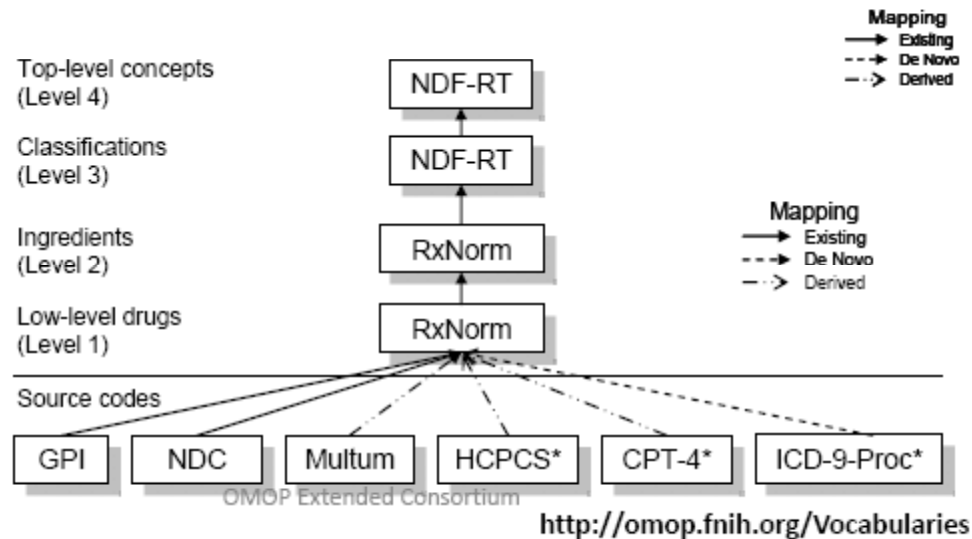
# Ontologies



## Standardizing conditions:



## Standardizing drugs:



- Sustainable processes and community support are required to maintain the mappings between the various medical vocabularies encountered across a network of active surveillance databases
  - OMOP has developed specifications for implementation of standard vocabularies for observational data analysis
  - Since vocabularies are ever changing, frequent updates to mappings are required
  - Perception of breaking new ground in using vocabularies in observational analysis
  - Expertise is very hard to find

- *OMOP Organizational Overview*
- *OMOP Areas of Focus*
  - *Data Partners and Simulated Data*
    - *Data Characteristics, Data Quality*
    - *Common Data Model*
    - *Ontologies*
  - **Cohort Selection**
  - Health Outcomes of Interest
  - Analytic Methods
- **Summary**

- Can we:
  - Develop a **standard procedure** to select patient cohorts / subgroups?
  - Apply that procedure to **disparate observational databases** (rapidly)?
  - **Compare** results returned by alternative definitions and against disparate databases?

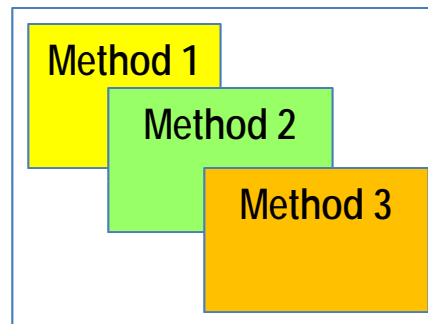
# Regularized Identification of Cohorts (aka RICO)



## Clinical Definitions

- Drug Cohorts
- Patient Subgroups
- Health Outcomes of Interest
- Etc...

## Analysis Methods Library



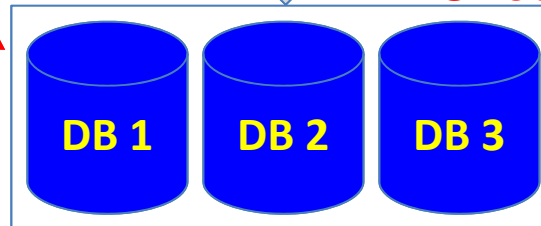
**Reused across multiple analyses**

### Definition Parameters

Aplastic Anemia
Ace inhibitors with hypertension
Females over 40
Elderly Alzheimer's
Pediatric bipolar

**Standardized Patient Selection**  
**RICO**

**Validated Once**



## Observational Databases in CDM Format

### Selected Patients (DB1)

Patient 0051...
Patient 7521
Etc....

### Selected Patients (DB2)

Patient 8001...
Patient 2902
Etc....

# RICO Parameter for Aplastic Anemia Definition #2



BEGIN

```
rico.insert_cohort(cohort_name=>'HOI Aplastic Anemia #2', cohort_concept_code=>'500000202',  
cohort_type=>'condition', first_exposure_only=>0, cohort_id=>my_cohort_id);
```

-- Broad Diagnosis SNOMED Codes

```
rico.insert_cohort_concept_ids(my_cohort_id,  
'137829,138723,140065,140681,4031699,4098025,4098026,4098027,4098028,4098145' ||  
,4098760,4100998,4100999,4101582,4101583,4120452,4120453,4125495,4125496,4125497' ||  
,4125498,4125499,4146086,4146087,4146088,4147492,4148471,4177177,4184200,4184758' ||  
,4184982,4186108,4187773,4188208,4211348,4211695,4225810,4228194,4234973,4298690' ||  
,4345236,4355763,4356365,4357375,4358305,4358788,4358817,4358844');
```

-- Broad Diagnosis MedDRA Codes

```
rico.insert_cohort_concept_ids(my_cohort_id,  
'35122796,35122805,37690284,37690338,37690643,37692638,37692921');
```

```
rico.insert_cohort_ruleset(cohort_id=> my_cohort_id,  
ruleset_name=>'HOI Aplastic Anemia #2 Bone Marrow Aspiration',  
ruleset_id=>my_ruleset_id);
```

```
rico.insert_cohort_rules(ruleset_id=>my_ruleset_id,  
concept_id_list=>  
'2002382,2002403,2108452,2108453,2212660,2212662,3045142,3048879,3635'  
days_before_list=>'60,60,60,60,60,60,60,60,60,60',  
days_duration_list=>'61,61,61,61,61,61,61,61,61,61');
```

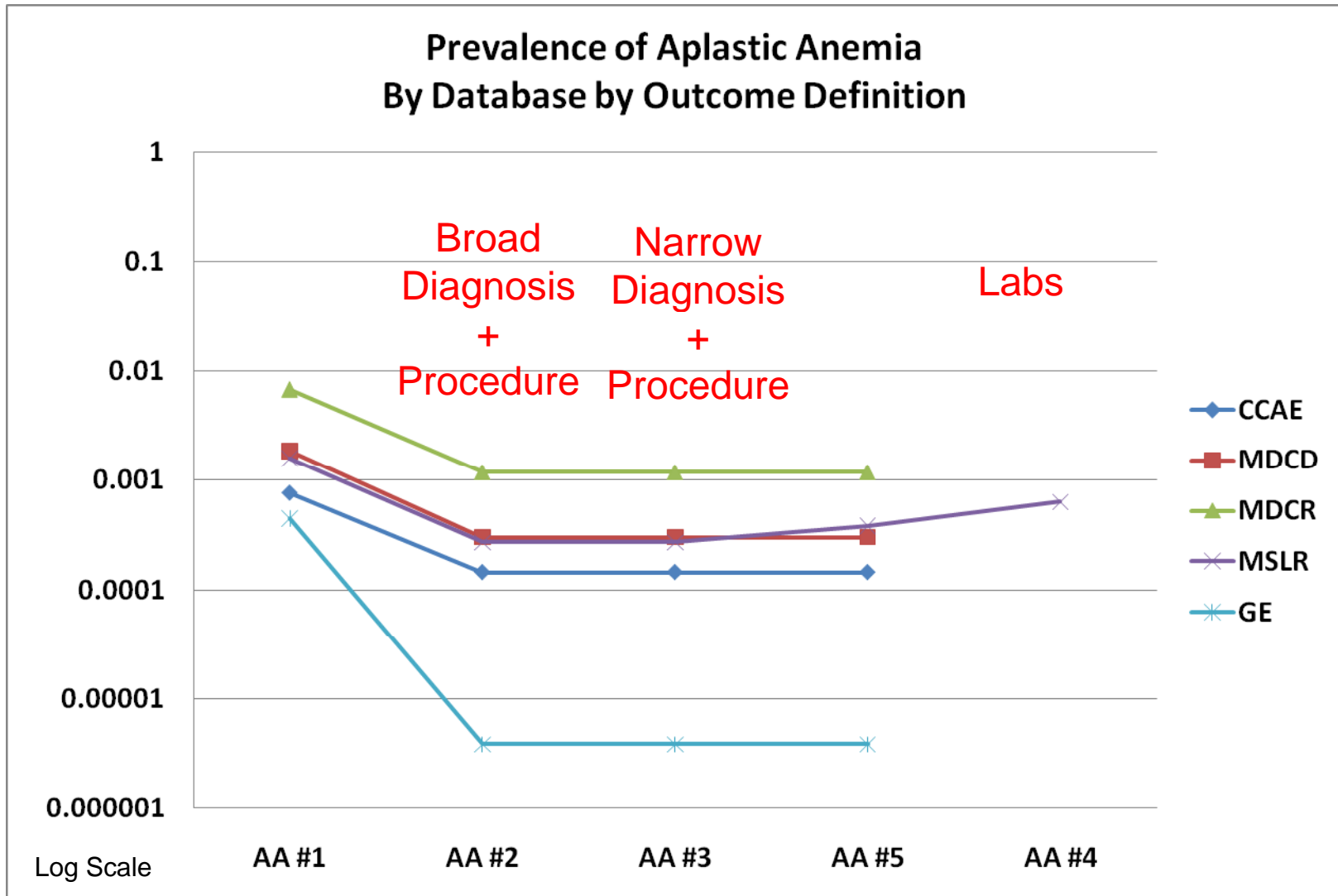
```
rico.create_cohort(my_cohort_id);
```

END;

#2

Occurrence of at least 1  
**Broad Diagnosis Code**  
AND at least one  
**Procedure Code** for bone  
marrow aspiration or  
biopsy w/in **60** days prior  
to diagnosis

# Aplastic Anemia Cohort Selection Results



- *OMOP Organizational Overview*
- *OMOP Areas of Focus*
  - *Data Partners and Simulated Data*
    - *Data Characteristics, Data Quality*
    - *Common Data Model*
    - *Ontologies*
  - *Cohort Selection*
  - **Health Outcomes of Interest**
  - *Analytic Methods*
- **Summary**

# 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

# Aplastic Anemia Alternative Definitions\*



**#1**

Occurrence of at least 1  
**Broad Diagnosis Code:**

284.0\*

284.8\*

284.9

**#2**

Occurrence of at least 1  
**Broad Diagnosis Code**  
AND at least one  
**Procedure Code** for bone  
marrow aspiration or  
biopsy w/in 60 days prior  
to diagnosis

**#3**

Occurrence of at least one  
**Narrow Diagnosis Code** –  
284.8\*, AND at least one  
**Procedure Code** for bone  
marrow aspiration or  
biopsy w/in 60 days prior  
to diagnosis

**#4**

Occurrence of at least  
two or three of the  
following **Lab Values:**  
WBC  $\leq 3.5 \times 10^9/L$   
Platelet Count  $\leq 50 \times 10^9/L$   
Hemoglobin  $\leq 100 \text{ g/L}$

**#5**

**# 3 OR #4**

- OSCAR provides a systematic approach for summarizing all data with the OMOP common data model
- Natural History Analysis (NATHAN) is an extension of OSCAR, where data characteristics can be produced for a particular subpopulation 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)
- Uses
  - Evaluate alternative cohort definitions (HOIs)
  - Comparisons between data sources
  - Providing context for interpreting and analyzing findings of drug safety studies

## Why Does this Matter to Industry?



- Develop **common library** of clinical definitions
  - Agreement among stakeholders, alternative definitions
- Create **technical implementation** for each definition
  - Patient selection independent of analysis programming and specific database
- **Re-use** standardized definitions **across analyses**
  - Transparent assumptions facilitate comparability and common interpretation
- **Evaluate suitability** of databases and definitions **prior to analysis**
  - Rapid sensitivity analyses

# Conceptual Example



## Selection and Restriction

- Age: **Persons Over 50** ✓
- Primary Selection: **Ace Inhibitor** ✓
- Additional Criteria
  - **First Exposure** ✓
  - **365 Days of Medical History** ✓

## Inclusion / Exclusion Criteria

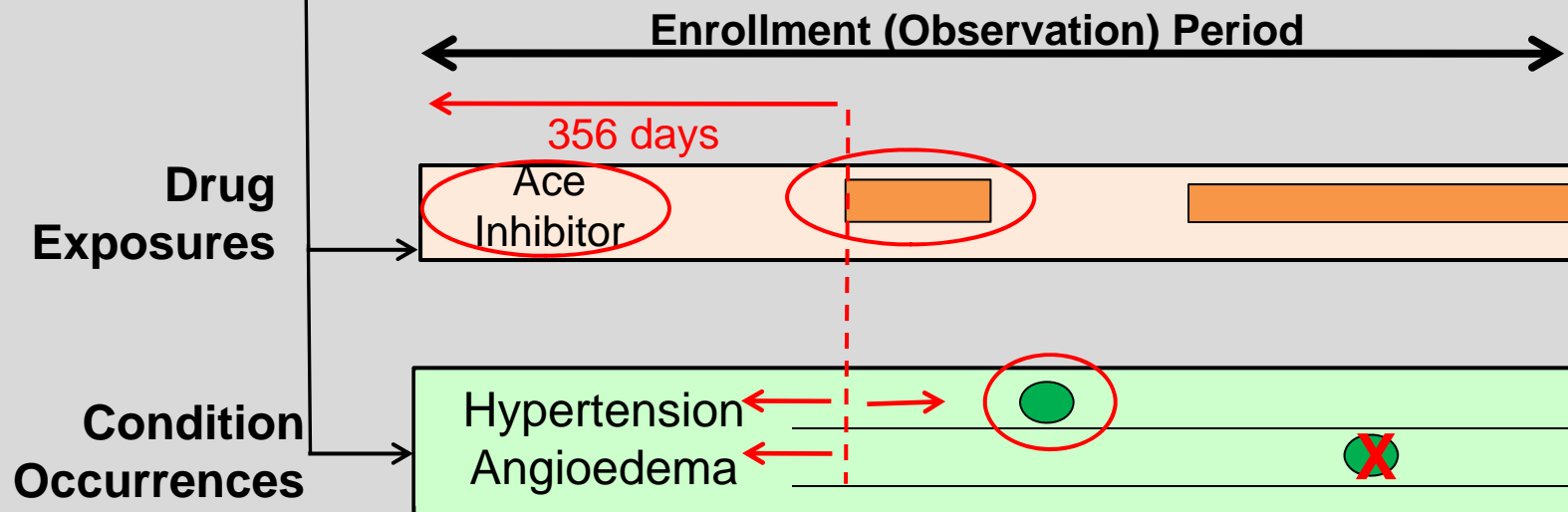
- Include: **Diagnosis of hypertension anytime** ✓
- Exclude: **Prior diagnosis of angioedema** ✓

CDM Format Patient Record

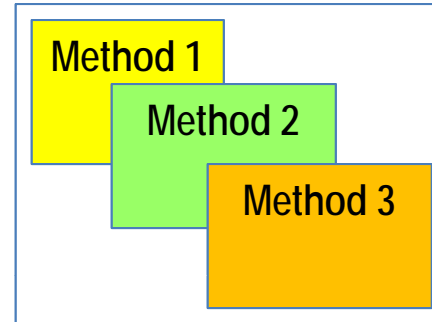
### Person

Patient ID: 3055  
Gender: Female  
Race: White  
Year of Birth: 1950 ←

**Patient 3055 would be included in the cohort**



## Analysis Methods Library



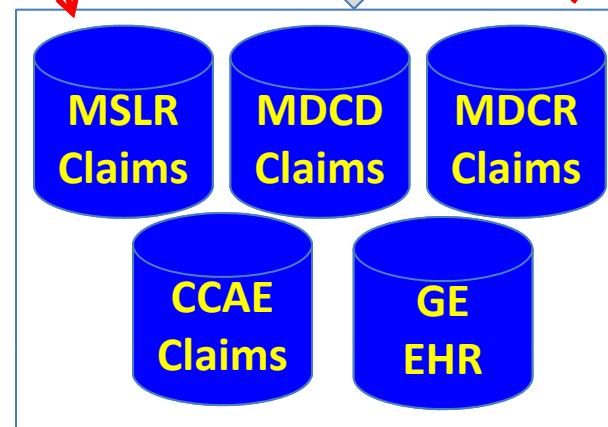
**Calculate Prevalence**

OMOP Health Outcomes of Interest Definitions
Angioedema (1)
<b>Acute Liver Injury (6)</b>
<b>Acute Renal Failure (3)</b>
Acute Myocardial Infarction (3)
<b>Aplastic Anemia (5)</b>
Bleeding (2)
Hip Fracture (4)
Hospitalization (1)
Mortality after Acute MI (4)

**Standardized Patient Selection**

**RICO**

**Selected Patients**



## Observational Databases

- Sustainable processes and community support are needed to expand the library of available Health Outcomes of Interest definitions
  - A consensus based approach similar to standards development is the first step
  - Current practice of using literature and manually revisiting source records is not scalable
  - Utilization of automation techniques to create HOI definitions should be explored

- *OMOP Organizational Overview*
- *OMOP Areas of Focus*
  - *Data Partners and Simulated Data*
    - *Data Characteristics, Data Quality*
    - *Common Data Model*
    - *Ontologies*
  - *Cohort Selection*
  - *Health Outcomes of Interest*
  - **Analytic Methods**
- **Summary**

- Exposure definition
  - Incident vs. prevalent exposure
  - Source of data capture
- Outcome definition
  - Incident vs. prevalent events
  - Diagnosis codes vs. HOI
- 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
  - Multivariate modeling
- Output metric/statistic
  - Estimation vs. testing
  - Relative vs. attributable risk
  - Measure of uncertainty

***Each method has input parameters that encode these choices***

## Disproportionality Analysis

	AE j = Yes	AE j = No
Drug i = Yes	a=20	b=100
Drug i = No	c=100	d=1080

- Distinct Patients
  - SRS
  - Modified SRS
- X
- MGPS  
BCPNN  
PRR  
Chi  
etc.
- X Stratified

- Temporal Pattern Discovery (WHO)

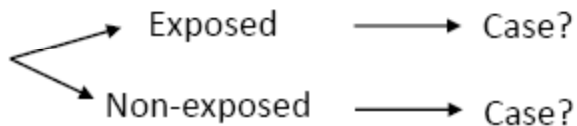
## Sequential Methods

	AE j = Yes	AE j = No
Drug i = Yes	a=20	
Drug i = No		

← Compare to baseline Poisson

- Maximized Sequential Probability Ratio Test (MaxSPRT)
- Conditional Sequential Sampling Procedure (CSSP)

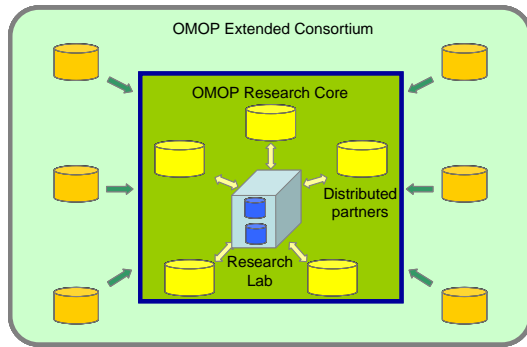
## Exposure Based Methods



- Observational screening
- HSIU
- Incident User Designs
- High-Dimensional Propensity Scoring
- Local control

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

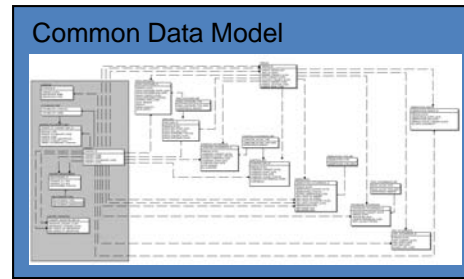
# Methods experiment workflow



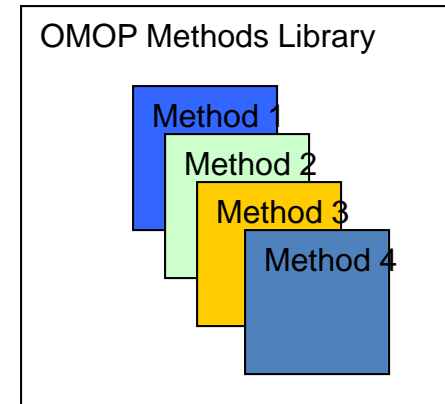
Testing in each source:  
 -accumulating over time  
 -against the entire dataset



- 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

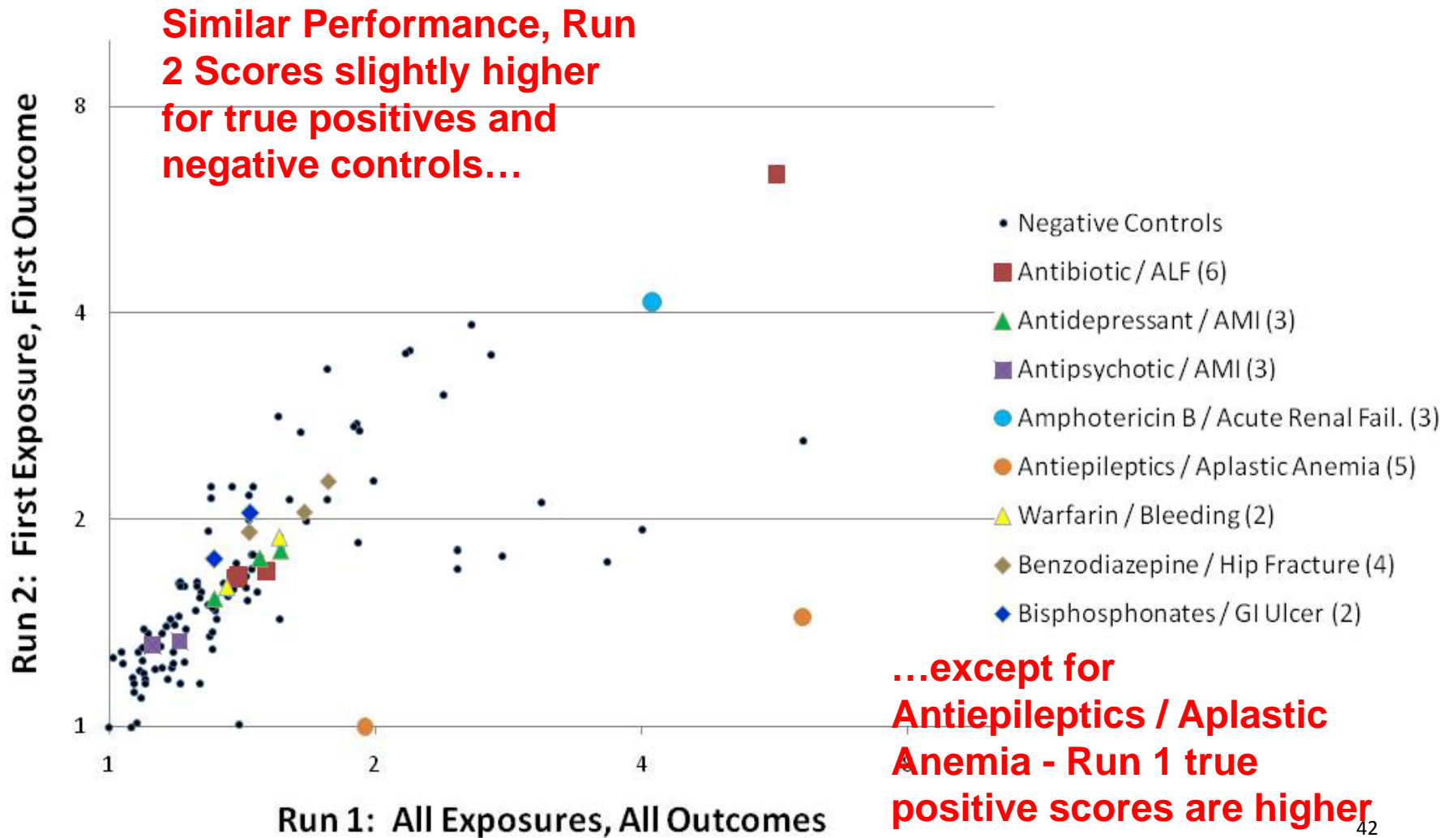


- Non-specified conditions**
- All outcomes in condition terminology
  - 'Labeled events' as reference
  - Warning
  - Precautions
  - Adverse Reactions
  - Postmarketing Experience

# Example of Methodological Research: Observational Screening for Selected HOIs



## Observational Screening Scores

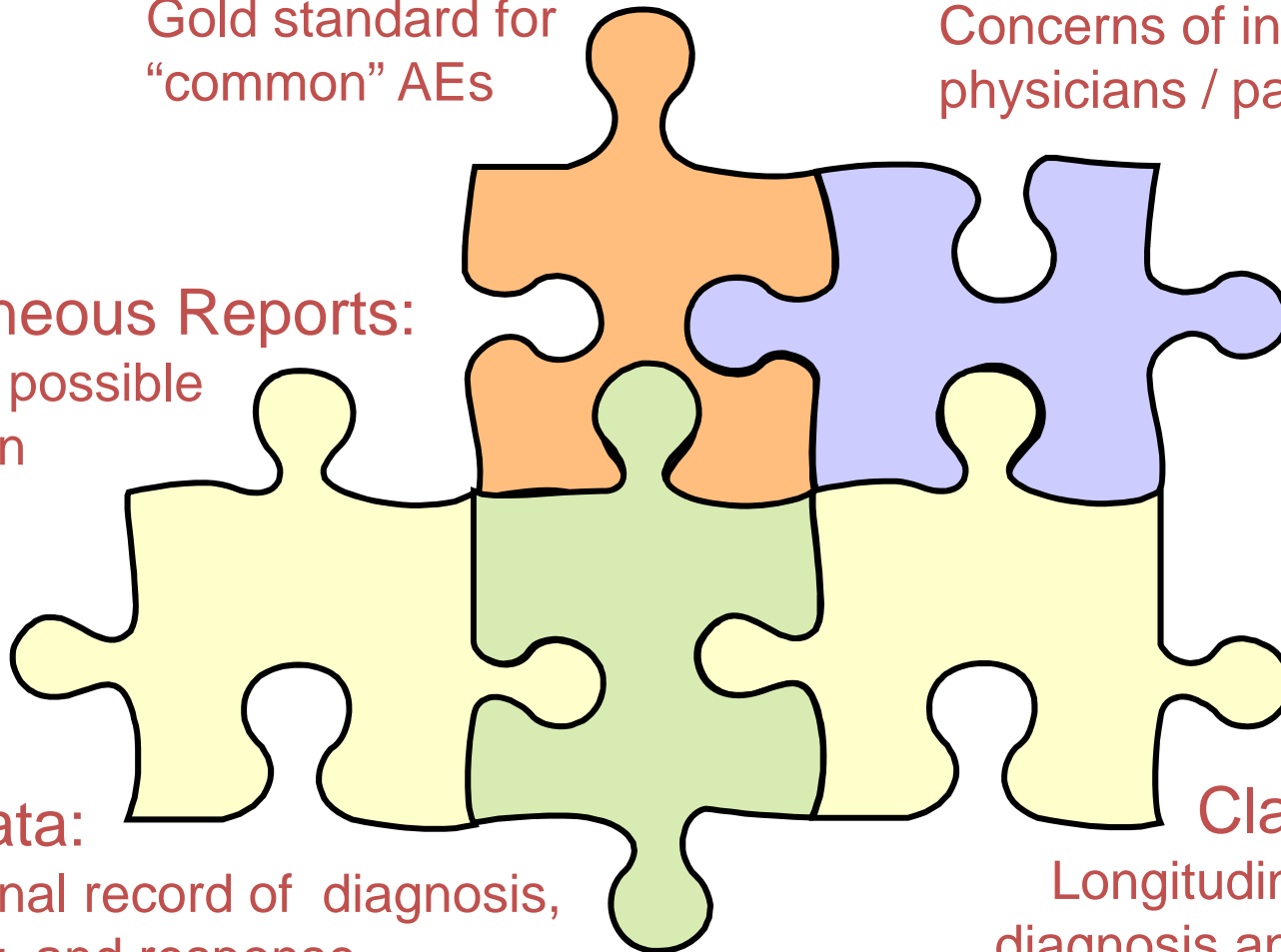


- *OMOP Organizational Overview*
- *OMOP Areas of Focus*
  - *Data Partners and Simulated Data*
    - *Data Characteristics, Data Quality*
    - *Common Data Model*
    - *Ontologies*
  - *Cohort Selection*
  - *Health Outcomes of Interest*
  - *Analytic Methods*
- **Summary**

**Clinical Trials:**  
Gold standard for  
“common” AEs

**Case Reports:**  
Concerns of individual  
physicians / patients

**Spontaneous Reports:**  
Broadest possible  
population



**EHR Data:**  
Longitudinal record of diagnosis,  
treatment, and response

**Claims Data:**  
Longitudinal record of  
diagnosis and treatment

***OMOP's research community requires active participation from all key stakeholders, including government, academia, industry, health care organizations, and patient groups.***

- **Governance**
  - 10 Executive Board members, chaired by FDA and managed by Foundation for NIH
  - 21 Advisory Board members
  - Led by 6 research investigators and Program Management Office
- **Methods**
  - 17 methods collaborators
- **Data**
  - **5** active distributed partners
  - 5 central databases included in the OMOP Research Lab
  - Simulated, claims and EHR datasets
- **Technology**
  - Secure virtual research lab
  - 2 data access models
  - 6 different systems architectures

***Over 100 researchers involved***

- Common Data Model
  - <http://omop.fnih.org/CDMandTerminologies>
- Standardized Technology
  - <http://omop.fnih.org/Vocabularies>
- Data Characteristics Tools
  - OSCAR – <http://omop.fnih.org/OSCAR>
  - GROUCH – <http://omop.fnih.org/GROUCH>
  - NATHAN – <http://omop.fnih.org/NATHAN>
- Health Outcomes of Interest Library
  - <http://omop.fnih.org/HOI>
- Methods Library
  - <http://omop.fnih.org/MethodsLibrary>
- Simulated Data
  - <http://omop.fnih.org/OSIM>

# Discussion