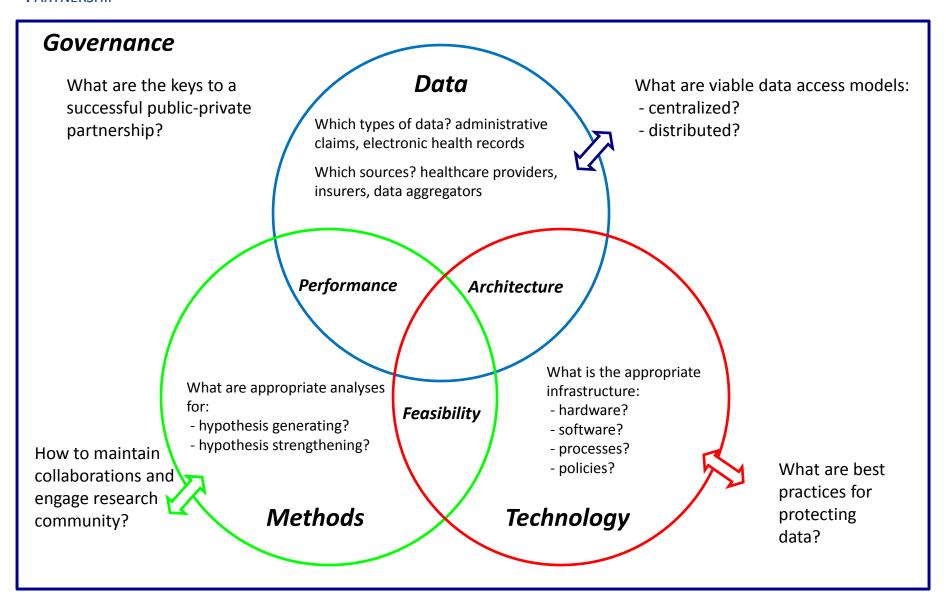
Visualization opportunities for systematic analyses of observational healthcare data: lessons from the Observational Medical Outcomes Partnership

Patrick Ryan on behalf of OMOP Research Team October 5, 2010

Outstanding questions for active surveillance



Observational Medical Outcomes Partnership

Established to inform the appropriate use of observational healthcare databases for active surveillance by:

- Conducting methodological research to empirically evaluate the performance of alternative methods on their ability to identify true drug safety issues
- Developing tools and capabilities for transforming, characterizing, and analyzing disparate data sources
- Establishing a shared resource so that the broader research community can collaboratively advance the science

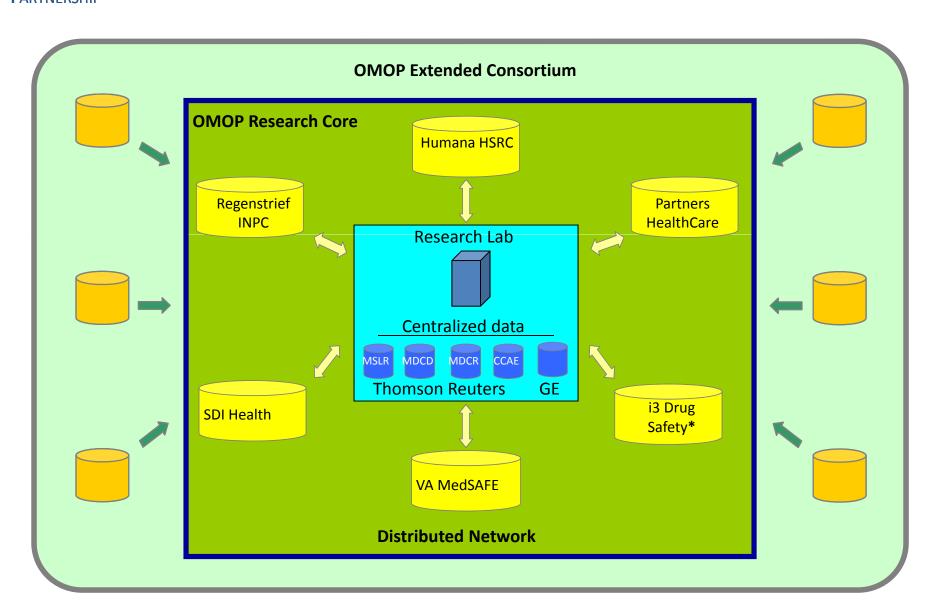
Partnership Stakeholders

A public-private partnership between industry, FDA and FNIH.

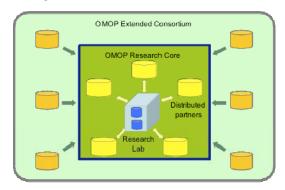
Stakeholder Groups

- FDA Executive Board [chair], Advisory Boards, PI
- Industry Executive and Advisory Boards, two PIs
- FNIH Partnership and Project Management, Research Core Staffing
- Academic Centers & Healthcare Providers Executive and Advisory Boards, three PIs, Distributed Research Partners, Methods Collaborators
- Database Owners Executive Board, Advisory Board, Pl
- Consumer and Patient Advocacy Organizations Executive and Advisory Board
- **US Veterans Administration** Distributed research partner

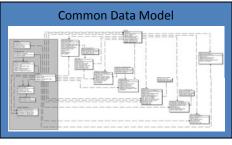
OMOP Data Community

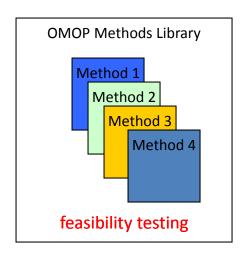


OMOP research experiment workflow











Health Outcomes of Interest

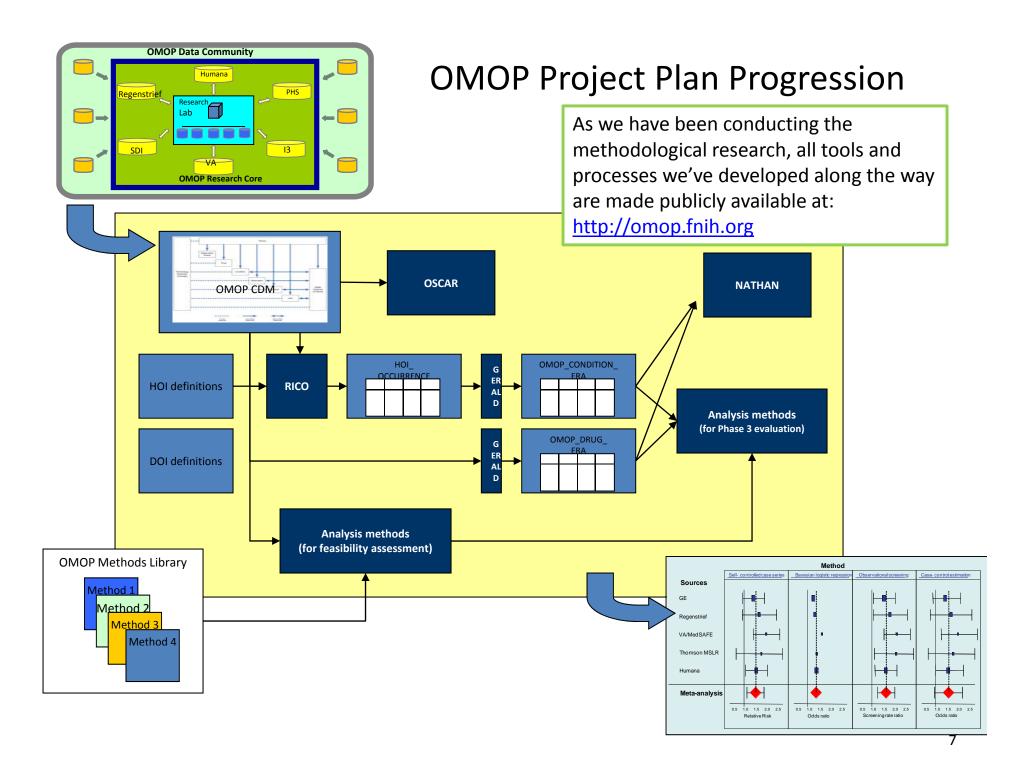
- Angioedema
- Aplastic Anemia
- Acute Liver Injury
- Bleeding
- GI Ulcer Hospitalization
- Hip Fracture
- Hospitalization
- Myocardial Infarction
- Mortality after MI
- Renal Failure

Drugs

- ACE Inhibitors
- Amphotericin B
- Antibiotics
- Antiepileptics
- Benzodiazepines
- Beta blockers
- Bisphosphonates
- Tricyclic antidepressants
- Typical antipsychotics
- Warfarin

Non-specified conditions

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



Observational Source Characteristics Analysis Report (OSCAR)

- Provides a systematic approach for summarizing observational healthcare data stored in the OMOP common data model
- Creates a structured output dataset of summary statistics of each table and field in the CDM
 - Categorical variables: one-, two-, and three-way stratified counts (e.g. number of persons with each condition by gender)
 - Continuous variables: distribution characteristics: min, mean, median, stdev, max, 25/75 percentile (e.g. observation period length)
 - OSCAR summaries from each source can be brought together to do comparative analyses

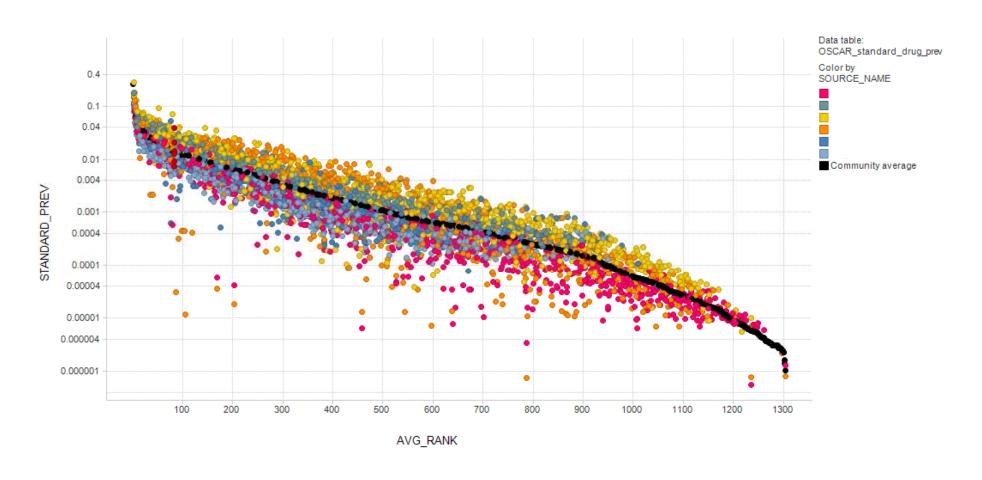
Uses

- 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 a particular drug or people with a specific condition)
- Providing context for interpreting and analyzing findings of drug safety studies

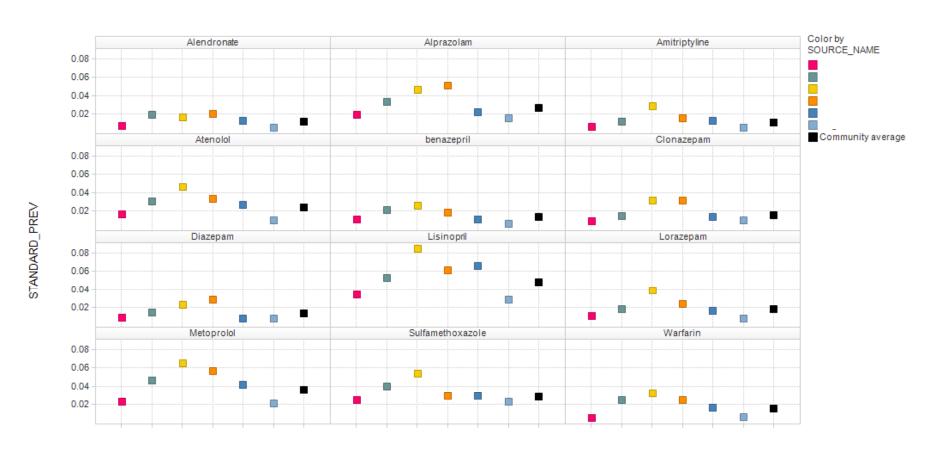
Characterization example: Drug prevalence

- Context: Typically we evaluate one drug at a time, against one database at a time
- In active surveillance, need to have ability to explore any medical product, across a network of disparate data sources
- Exploration: what is the prevalence of all medical products across all data sources?
 - Crude prevalence: # of persons with at least one exposure / # of persons
 - Standardized prevalence: Adjusted by age and gender to US Census
 - Age-by-gender Strata-specific prevalence
- Not attempting to get precise measure of exposure rates for a given product, but instead trying to understand general patterns across data community for all medical products

Standardized drug prevalence (age*gender stratified annualized rates; standardized to US Census)

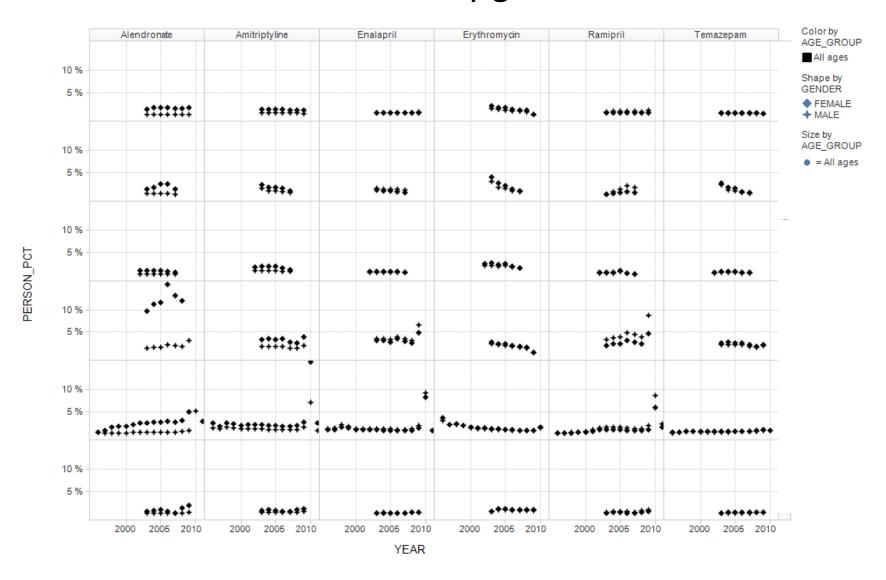


Standardized prevalence for select drugs



SOURCE_NAME

Source-specific drug prevalence, by year by age by gender

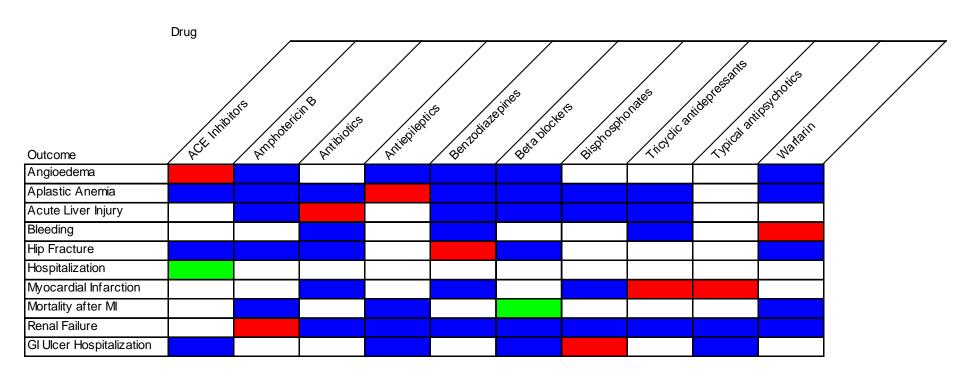


What methods are most appropriate for signal refinement? Multiple alternative approaches identified that deserve empirical testing to measure performance

Method name	Parameter combinations	Release
Disproportionality analysis (DP)	112	
Univariate self-controlled case series (USCCS)	64	2-Apr-10
Observational screening (OS)	162	8-Apr-10
Multi-set case control estimation (MSCCE)	32	16-Apr-10
Bayesian logistic regression (BLR)	24	21-Apr-10
Case-control surveillance (CCS)	48	2-May-10
IC Temporal Pattern Discovery (ICTPD)	84	23-May-10
Case-crossover (CCO)	48	1-Jun-10
HSIU cohort method (HSIU)	6	8-Jun-10
Maximized Sequential Probability Ratio Test (MSPRT)	144	25-Jul-10
High-dimensional propensity score (HDPS)	144	6-Aug-10
Conditional sequential sampling procedure (CSSP)	144	30-Aug-10
Statistical relational learning (SRL)		
Incident user design (IUD-HOI)		

http://omop.fnih.org/MethodsLibrary

'Ground truth' for Monitoring Health Outcomes of Interest



Legend Total

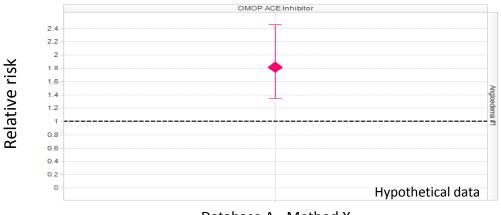
B- 'True positive' benefit 2

R- 'True positive' risk 9

N- 'Negative control' 44

Studying method performance for 'signal refinement'

- Apply method with specific set of parameter settings to a database for a drug-outcome pair that has a prior suspicion of being potentially related
- Example: Run method X on database A for ACE inhibitors Angioedema

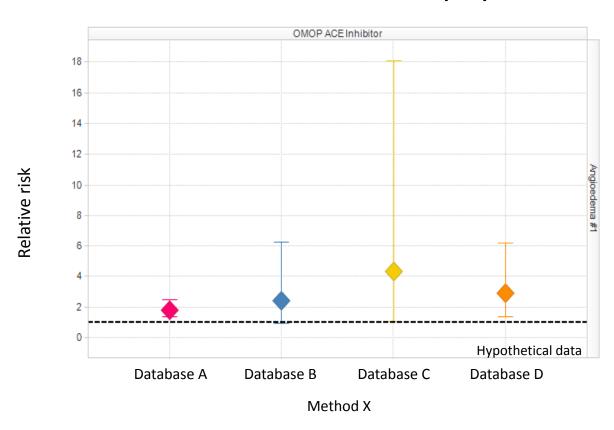


Database A - Method X

Challenges:

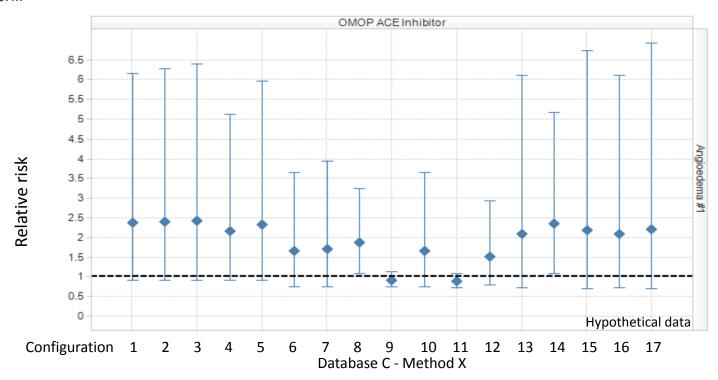
- How to put the resulting score in context?
- What if we modified one of the methods parameters?
- What if we applied the method to a different database?
- If we had applied the same method to other drug-outcome pairs, what types of scores would we expect?
 - How many other true positives would get a RR > 1.8?
 - How many false positives would be identified with a threshold of 1.8?

How do effect estimates vary by database?



- Each database may have unique source population characteristics that can influence method behavior, including:
 - Sample size
 - Length and type of longitudinal data capture
 - Population demographics, such as age, gender
 - Disease severity, including comorbidities, concomitant medications and health service utilization patterns

How do estimates vary by method parameter settings?



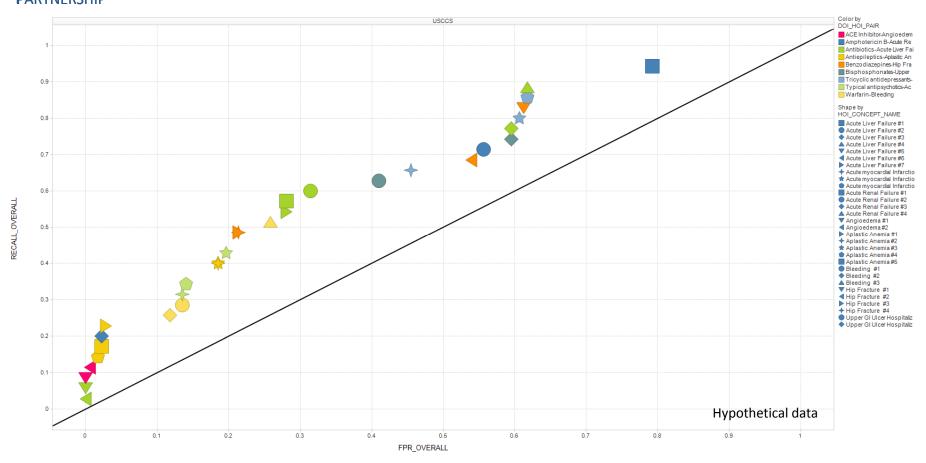
- Performance can be sensitive to various factors, including:
 - Length of washout period to identify incident use
 - Definition of time-at-risk
 - Choice of comparator
 - Number and types of covariates to include in propensity score modeling
 - Statistical approach for adjustment: matching vs. stratification vs. multivariate modeling
- 'Optimal' settings may vary by database and/or the drug-outcome pair in question

How does method perform against other 'benchmark' true positives and negative controls?



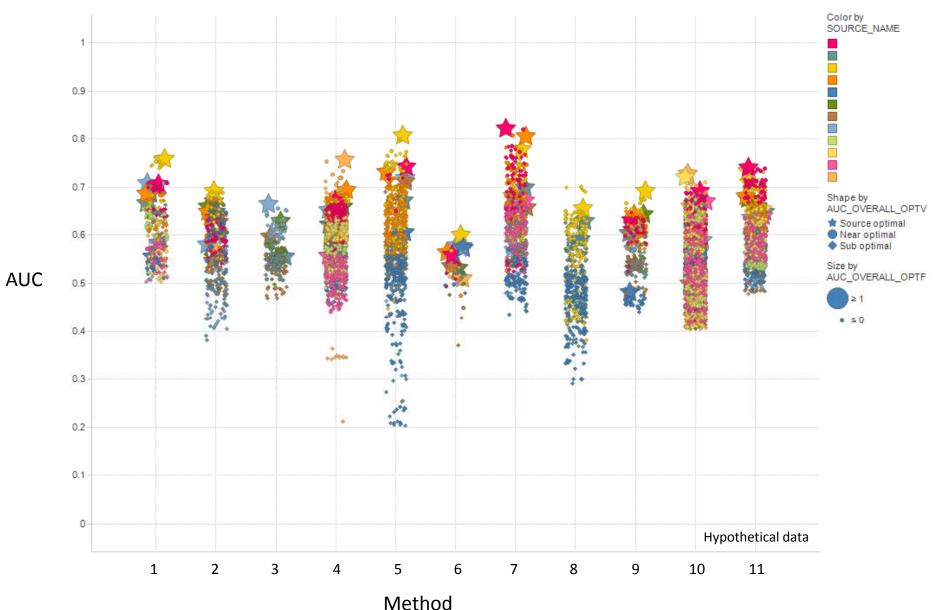
 It is important to establish operating characteristics of any method, when applied across a network of databases, as part of the 'validation before adoption for signal refinement'

Receiver Operating Characteristic (ROC) curve



- ROC plots sensitivity (recall) vs. false positive rate (FPR)
- Area under ROC curve (AUC) provides probability that method will score a randomly chosen true positive drug-outcome pair higher than a random unrelated drug-outcome pair
- AUC=1 is perfect predictive model, AUC=0.50 is random guessing (diagonal line)

Visualizing performance of alternative methods across a network of databases



Opportunities for exploratory visualization

- The use of exploratory visualization has been an invaluable component of the OMOP research team's learning process.
- We have made extensive use of Spotfire® as a tool to study the data characteristics of each participating source, and evaluate the method performance across the community.
- Visualization tools have enabled interactive exploration of OMOP's summary results, and have provided a consistent framework for effectively communicating with our distributed partners and other research collaborators.
- Exploratory visualization may become an increasingly valuable tool for using observational healthcare data to gain a greater understanding of the effects of medical products.

Contact information

Patrick Ryan Research Investigator ryan@omop.org

Thomas Scarnecchia
Executive Director
scarnecchia@omop.org

Emily Welebob Senior Program Manager, Research welebob@omop.org

OMOP website: http://omop.fnih.org