A Few Words About Observational Data

- **Very large** datasets: millions of lives
  - **Claims**: represent a financial transaction and include many biases and ‘errors’
  - **EHR**: represent a ‘clinical’ record mostly but are often incomplete; Rx written not filled
- Reflect **underlying health care delivery** system
- **Non-randomized**: measureable and un-measureable confounders and biases
- From Pharma company: ‘exploring’ database has strong **Regulatory/Criminal repercussions**
Working with observational databases really entails:

- **Exposure**
  - Prescriptions written
  - Prescriptions filled
    - How were they taken?
    - What about prn use?

- **Outcome**
  - Diagnosis codes alone
  - Dx + procedure?
  - Dx, procedure, lab results?
  - Site of care?
  - Death?
Observational Data: Information Asymmetry

- Many ‘benefits’ (improvement in signs/symptoms, ADL, QoL) are not ‘clinical diagnoses’ so they are not captured
  - Limited capture of utilization-based measures ("switching drugs", change in ER/hospitalization) or reduction in clinical events
- Most ‘risks’ are clinical and would be captured in clinical encounter
  - But we do not know how impactful they are nor what perception is by patients and providers

Considerations in Clinical Information

By Perspective

From Stang et al., Am J Therap, 2008
Outstanding questions for active surveillance

**Governance**
- What are the keys to a successful public-private partnership?

**Data**
- Which types of data? administrative claims, electronic health records
- Which sources? healthcare providers, insurers, data aggregators

**Performance**
- What are appropriate analyses for:
  - hypothesis generating?
  - hypothesis strengthening?

**Architecture**
- What is the appropriate infrastructure:
  - hardware?
  - software?
  - processes?
  - policies?

**Feasibility**
- What are viable data access models:
  - centralized?
  - distributed?

**Technology**
- What are best practices for protecting data?

**Methods**
- How to maintain collaborations and engage research community?

---

Breadth and diversity of OMOP research community

**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 5 research investigators and PMO

- **Methods**
  - 17 methods collaborators

- **Data**
  - 6 distributed partners
  - 5 central databases included in the OMOP Research Lab

- **Technology**
  - 2 data access models, 7 different systems architectures

**Over 100 partners collaborating to advance the science of drug safety!**
### Executive Board

A multi-stakeholder group, the OMOP Executive Board oversees the operation of the Partnership.

<table>
<thead>
<tr>
<th>Name</th>
<th>Position and Affiliations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Janet Woodcock, MD</td>
<td>Director, Center for Drug Evaluation and Research, Food and Drug Administration Chair, Observational Medical Outcomes Partnership Executive Board</td>
</tr>
<tr>
<td>Rebecca Burkholder</td>
<td>Vice President of Health Policy, The National Consumers League</td>
</tr>
<tr>
<td>Sherine Gabriel, MD, MSc</td>
<td>Professor of Medicine and Epidemiology, The Mayo Clinic</td>
</tr>
<tr>
<td>Cynthia Gilman, JD</td>
<td>Special Assistant to the President for Advancement of Cancer Research and Collaborative Partnerships, Henry Jackson Foundation</td>
</tr>
<tr>
<td>Jesse L. Goodman, MD, MPH</td>
<td>Chief Scientist and Deputy Commissioner for Science and Public Health (acting), Food and Drug Administration</td>
</tr>
<tr>
<td>Ronald L. Krall, MD</td>
<td>Former Senior Vice President and Chief Medical Officer, GlaxoSmithKline</td>
</tr>
<tr>
<td>Richard Platt, MD, MSc</td>
<td>Professor and Chair of the Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care</td>
</tr>
<tr>
<td>Stephen Spielberg, MD, PhD</td>
<td>Marion Merrell Dow Chair in Pediatric Pharmacogenomics, Children’s Mercy Hospital and Dean Emeritus, Dartmouth Medical School</td>
</tr>
<tr>
<td>Brian Strom, MD, MPH</td>
<td>George S. Pepper Professor of Public Health and Preventive Medicine, Professor of Biostatistics and Epidemiology, Medicine, and Pharmacology; Chair, Department of Biostatistics and Epidemiology; Director, Center for Clinical Epidemiology and Biostatistics; Vice Dean for Institutional Affairs, University of Pennsylvania School of Medicine Senior Advisor to the Provost for Global Health Initiatives, University of Pennsylvania</td>
</tr>
<tr>
<td>David Wheadon, MD</td>
<td>Senior Vice President, Pharmaceutical Research and Manufacturers of America (PhRMA)</td>
</tr>
</tbody>
</table>

### Research Investigators

The Principal Investigators (PIs) are the lead scientists for the OMOP project and guide and participate in the research across all four project phases.

- **Marc Overhage, MD, PhD:** Director, Medical Informatics and Research Scientist, Regenstrief Institute, Inc.; Regenstrief Professor of Medical Informatics, Indiana University School of Medicine, CEO; President of the Indiana Health Information Exchange
- **Paul Stang, PhD:** Senior Director, Epidemiology, Johnson & Johnson Pharmaceutical Research and Development
- **Abraham G. Hartzema PharmD, MSPH, PhD:** Professor and Eminent Scholar, Pharmaceutical Outcomes & Policy, Perry A. Foote Chair in Health Outcomes Research, University of Florida College of Pharmacy
- **Judy Racoosin, MD, MPH:** Sentinel Initiative Scientific Lead, US Food and Drug Administration
- **Patrick Ryan:** Manager Drug Development Sciences, GlaxoSmithKline R&D OMOP Co-Investigator
OMOP’s Methods To Date

- Disproportionality analysis (DP)
- Observational screening (OS)
- Univariate self-controlled case series (USCCS)
- Case-control surveillance (CCS)
- Bayesian logistic regression (BLR)
- Multi-set case control estimation (MSCCE)
- Maximized sequential probability ratio test (MaxSPRT)
- IC Temporal Pattern Discovery (ICTPD)
- High-dimensional propensity score (HDPS)
- Conditional sequential sampling procedure (CSSP)
- Case-crossover (CCO)
- HSIU cohort method (HSIU)
- Statistical relational learning (SRL)
- Incident user design (IUD)
- Multivariate self-controlled case series
- Case-time control
- Lasso propensity scoring
- Online algorithms
- OMOP Cup (50+ submissions)

Methodological considerations common across multiple approaches

- 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 user input parameters that encode these choices
Analysis problems under study by OMOP

- **Monitoring of Health Outcomes of Interest (HOIs):**
  - Estimate the strength of the association between drug exposure and specific events (e.g., acute liver failure, bleeding, MI)
  - Modest in number so can customize analytic approach
  - Expert assessment of drug-HOI causal associations based on literature search

- **Identification of non-specified associations:**
  - More exploratory in nature
  - Same goal: estimate the strength of the association between drug exposure and conditions
  - Necessarily more generic analyses (e.g., adjust for age and sex)
  - Causality assessment relies on the product labels

- **Performance against simulated data**
  - Complement ‘real world’ experiments
  - Ground truth explicitly defined

**SAB/HiAB Review Process:** July 2009 Methods strategy / briefing web meeting
OMOP Project Plan Progression:
Tools to support observational database research

OMOP Data Community
OMOP Data Community

OMOP Research Core
OMOP Research Core

OMOP Methods Library
OMOP Methods Library

OMOP Data Model
OMOP Data Model

OMOP Methods Library
OMOP Methods Library

OMOP Data Model
OMOP Data Model

OMOP Confidential
OMOP Confidential

Developed OMOP Common Data Model

Each data source has been successfully transformed to the OMOP common data model

OMOP Confidential
OMOP Confidential
Establishing a Common Data Model

- Developed with broad stakeholder input
- Designed to accommodate disparate types of data (claims & EHRs)
- Applied successfully across OMOP data community

http://omop.fnih.org/CDMandTerminologies

OMOP Confidential 17

Standardizing terminologies to accommodate disparate observational data sources

http://omop.fnih.org/Vocabularies

OMOP Confidential 18
### Executed OSCAR

**Observational Source Characteristics Analysis Report (OSCAR)** - provides a systematic approach for summarizing all data within the OMOP common data model.

### 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.

[http://omop.fnih.org/OSCAR](http://omop.fnih.org/OSCAR)
Initiated Health Outcomes of Interest (HOI) Library

Execute Generalized ERA Logic Developer (GERALD) to populate the HOI ERA table

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

Defined OMOP Drugs of Interest (DOI)

Drugs
- ACE Inhibitors
- Amphotericin B
- Antibiotics
- Antiepileptics
- Benzodiazepines
- Beta blockers
- Bisphosphonates
- Tricyclic antidepressants
- Typical antipsychotics
- Warfarin
Natural History Analysis (NATHAN)

- OSCAR provides a systematic approach for summarizing all data within 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)
- Additional NATHAN summary statistics provide temporal assessment, relative to index date
  - Ex. conditions 30d prior to drug start
  - Ex. drug exposure any time prior to incident condition
- Uses:
  - Evaluate alternative cohort definitions (HOIs)
  - Comparisons between data sources
  - Providing context for interpreting and analyzing findings of drug safety studies

http://omop.fnih.org/NATHAN
OMOP’s Methods Landscape

**Disproportionality Analysis**

<table>
<thead>
<tr>
<th>Drug i = Yes</th>
<th>AE j = Yes</th>
<th>AE j = No</th>
</tr>
</thead>
<tbody>
<tr>
<td>c=100</td>
<td>d=1080</td>
<td></td>
</tr>
</tbody>
</table>

- Distinct Patients
- SRS
- Modified SRS
- MGPS
- BCPNN
- Stratified
- Temporal Pattern Discovery (WHO)

**Sequential Methods**

<table>
<thead>
<tr>
<th>Drug i = Yes</th>
<th>AE j = Yes</th>
<th>AE j = No</th>
</tr>
</thead>
<tbody>
<tr>
<td>α=20</td>
<td>d=100</td>
<td></td>
</tr>
</tbody>
</table>

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

**Exposure Based Methods**

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

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

Case Based Methods
- Exposed? → Case
- Exposed? → Non-case

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

Other Methods

• Hi-Dimensional logistic regression
• Statistical relational learning

Future Methods

• Multivariate self-controlled case series
• Case-time control
• Lasso propensity scoring
• Online algorithms
• OMOP Cup (50+ submissions)

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

Conduct Analyses to Evaluate Methods
OMOP Analysis Process:
Designed for active surveillance
Applicable to broader research applications

Derivative Products and Impacts

- Validation tools
- Standards: Connected to Office of the National Coordinator
- Feedback loop to data capture in EHRs
- Decision-making tools
- Visualizations
- ‘Natural Experiments’
For further information

http://omop.fnih.org