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**
Observational Medical Outcomes Partnership

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?

---

The News and Observer
Sunday, December 5, 1993

*Available archived by Free Library of Philadelphia: This document contains images. For more, see [the Philadelphia Inquirer's repository](https://archive.org/details/the_news_and_observer_1993-12-05).*
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**
What are viable data access models:
- centralized?
- distributed?

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

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

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

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

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

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

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!*

DO NOT DISTRIBUTE
Executive Board

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

Janet Woodcock, MD
Director, Center for Drug Evaluation and Research, Food and Drug Administration
Chair, Observational Medical Outcomes Partnership Executive Board

Rebecca Burkholder
Vice President of Health Policy, The National Consumers League

Sherine Gabriel, MD, MSc
Professor of Medicine and Epidemiology, The Mayo Clinic

Cynthia Gilman, JD
Special Assistant to the President for Advancement of Cancer Research and Collaborative Partnerships, Henry Jackson Foundation

Jesse L. Goodman, MD, MPH
Chief Scientist and Deputy Commissioner for Science and Public Health (acting), Food and Drug Administration

Ronald L. Krall, MD
Former Senior Vice President and Chief Medical Officer, GlaxoSmithKline

Richard Platt, MD, MSc
Professor and Chair of the Department of Ambulatory Care and Prevention, Harvard Medical School and Harvard Pilgrim Health Care

Stephen Spielberg, MD, PhD
Marion Merrell Dow Chair in Pediatric Pharmacogenomics, Children’s Mercy Hospital and Dean Emeritus, Dartmouth Medical School

Brian Strom, MD, MPH
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

David Wheaton, MD
Senior Vice President, Pharmaceutical Research and Manufacturers of America (PhRMA)

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

OMOP Project Plan Progression:
Tools to support observational database research

Developed OMOP Common Data Model

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

Standardizing terminologies to accommodate disparate observational data sources

Standardizing conditions:

- Top-level classifications (Level 3)
- Higher-level classifications (Level 2)
- Low-level concepts (Level 1)

Source codes: MedDRA, SNOMED-CT, ICD-9-CM, Read, Oxmis

http://omop.fnih.org/Vocabularies

Standardizing drugs:

- Top-level concepts (Level 4)
- Ingredients (Level 2)
- Low-level drugs (Level 1)

Source codes: GPI, NDC, Multum, HCPCS*, CPT-4*, ICD-9-Proc*

OMOP Confidential
**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.

**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.
Developed Methods Library

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

Conduct Analyses to Evaluate Methods

OMOP Confidential
For further information

http://omop.fnih.org