

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Developing Tools for Conducting
Observational Database Research
Across a Network of Data Sources

Paul Stang, PhD
on behalf of the OMOP team

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

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**


NUMBER: A01322 SEX: F AGE 47 ENTRY11/04/89 EVENT7 05/23/80

Plan	Site	Provider	ICD	CPT		
	1-OFFICE/CLIN	0106903	7169	ARTHROPATHY NOS	90015	VISIT
	1-OFFICE/CLIN	0106903	5990	URIN TRACT INFECTION NOS	81000	URINALYSIS WITH MICR
	1-OFFICE/CLIN	0106903	5990	URIN TRACT INFECTION NOS	90070	VISIT
	1-OFFICE/CLIN	0106903	9953	ALLERGY, UNSPECIFIED	90060	VISIT
	1-OFFICE/CLIN	0106903	8953	ALLERGY, UNSPECIFIED	J0420	INJECTION
E	1-OFFICE/CLIN	8608905	8751	CHOLECYSTITIS NEC	90070	VISIT
*	1-OFFICE/CLIN	1705176	7890	ABDOMINAL PAIN	90070	VISIT
*	1-OFFICE/CLIN	1705176	7890	ABDOMINAL PAIN	90630	VISIT
*	1-OFFICE/CLIN	2902146	789	OTH ABDOMEN/PELVIS SYMP	90017	VISIT
	3-LAB/RADIOL	3503411	V726	LABORATORY EXAMINATION	80019	19 OR MORE BLOOD/URI
*	3-LAB/RADIOL	3503411	V726	LABORATORY EXAMINATION	82150	ASSAY OF SERUM AMYLA
	3-LAB/RADIOL	3503411	V726	LABORATORY EXAMINATION	83545	AUTO-ASSAY SERUM IRO
	3-LAB/RADIOL	3503411	V726	LABORATORY EXAMINATION	83690	ASSAY BLOOD LIPASE
	3-LAB/RADIOL	3503411	V726	LABORATORY EXAMINATION	83720	BLOOD LIPOPROTEIN AS
	3-LAB/RADIOL	3503411	V726	LABORATORY EXAMINATION	85025	AUTOMATED HEMOGRAM
	3-LAB/RADIOL	3503411	V726	LABORATORY EXAMINATION	85651	RBC SEDIMENTATION RA
*	6-HOSPITAL OP	1606466	7890	ABDOMINAL PAIN	74246	CONTRAST XRAY UPPER
*	6-HOSPITAL OP	1606466	7890	ABDOMINAL PAIN	76700	ECHO EXAM OF ABDOMEN
*	6-HOSPITAL OP	5006031	V725	RADIOLOGICAL EXAM NEC	80005	X-RAYS
*	1-OFFICE/CLIN	2902146	532	DUODENAL ULCER	90050	VISIT
*	1-OFFICE/CLIN	2902146	786	OTH ABDOMEN/PELVIS SYMP	90060	VISIT
*	3-LAB/RADIOL	1106534	7890	ABDOMINAL PAIN	88305	TISSUE EXAM BY PATHO
*	3-LAB/RADIOL	1106534	7890	ABDOMINAL PAIN	88312	SPECIAL STAINS
*	3-LAB/RADIOL	1106534	7890	ABDOMINAL PAIN	89060	EXAM, SYNOVIAL FLUID
*	6-HOSPITAL OP	2902146	789	OTH ABDOMEN/PELVIS SYMP	43239	UPPER GI ENDOSCOPY,
*	9-IP SURGICTR	8806031	5355	GASTRITIS/DUODENITIS NOS	43234	UPPER GI ENDOSCOPY,
*	1-OFFICE/CLIN	2902146	532	DUODENAL ULCER	90050	VISIT
	1-OFFICE/CLIN	0106903	7865	CHEST PAIN	80019	19 OR MORE BLOOD/URI
	1-OFFICE/CLIN	0106903	7865	CHEST PAIN	83705	ASSAY BLOOD LIPID GR
	1-OFFICE/CLIN	0106903	7865	CHEST PAIN	84478	ASSAY BLOOD TRIGLYCE
	1-OFFICE/CLIN	0106903	7865	CHEST PAIN	85031	MANUAL HEMOGRAM,COMP
	1-OFFICE/CLIN	0106903	7865	CHEST PAIN	85651	RBC SEDIMENTATION RA
	1-OFFICE/CLIN	0106903	7865	CHEST PAIN	90060	VISIT
	1-OFFICE/CLIN	0106903	7865	CHEST PAIN	93000	ELECTROCARDIOGRAM, C

**OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP**

Working with observational medical outcomes entails

- **Exposure**
 - Prescriptions written
 - Prescriptions filled
 - How were they taken?
 - What about prn use?
- **Outcome**
 - Diagnosis codes alone
 - Dx + procedure?
 - Dx, procedure, lab res
 - Site of care?
 - Death?



The News and Observer Sunday, December 8, 1991

"An inhaler prescribed by her doctor helps Gail Pouncey's smoke-scarred lungs—like many survivors, she suffered respiratory injuries in the Imperial fire."

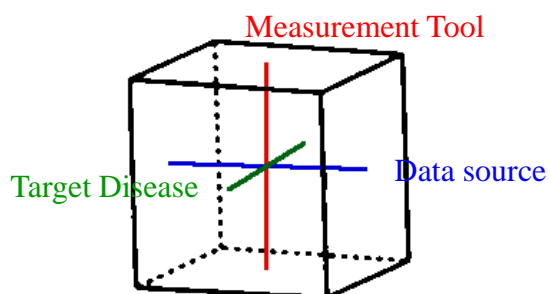
Observational Data: Information Asymmetry

- Many '**benefits**' (improvement in signs/syptoms, 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

5

Considerations in Clinical Information

By Perspective



From Stang et al., Am J Therap, 2008

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

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

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

Performance **Architecture**

Feasibility

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

How to maintain collaborations and engage research community?

Methods **Technology**

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

What are best practices for protecting data?

OMOP CONFIDENTIAL – Not for Distribution Page 7

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

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

Data

Methods **Technology**

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

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

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 Wheadon, MD

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

DO NOT DISTRIBUTE

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

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

DO NOT DISTRIBUTE

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

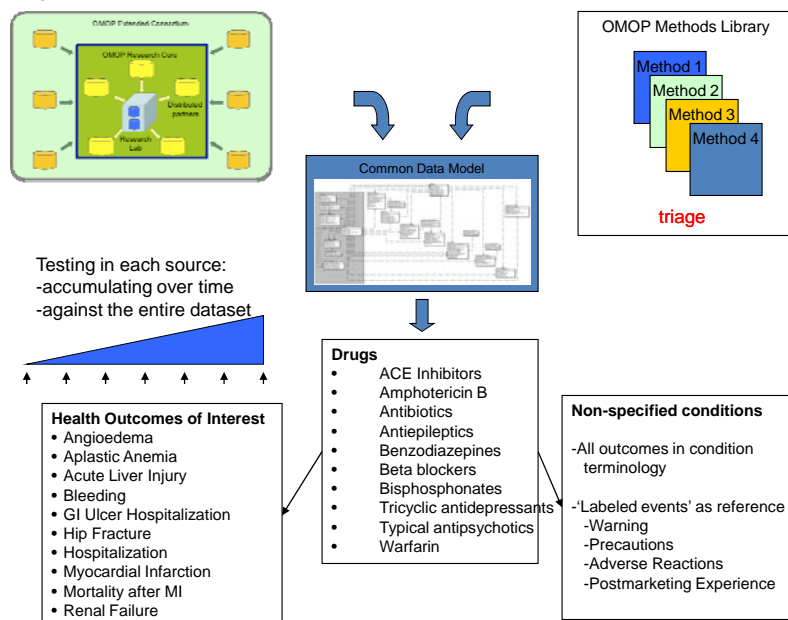
Each method has user input parameters that encode these choices

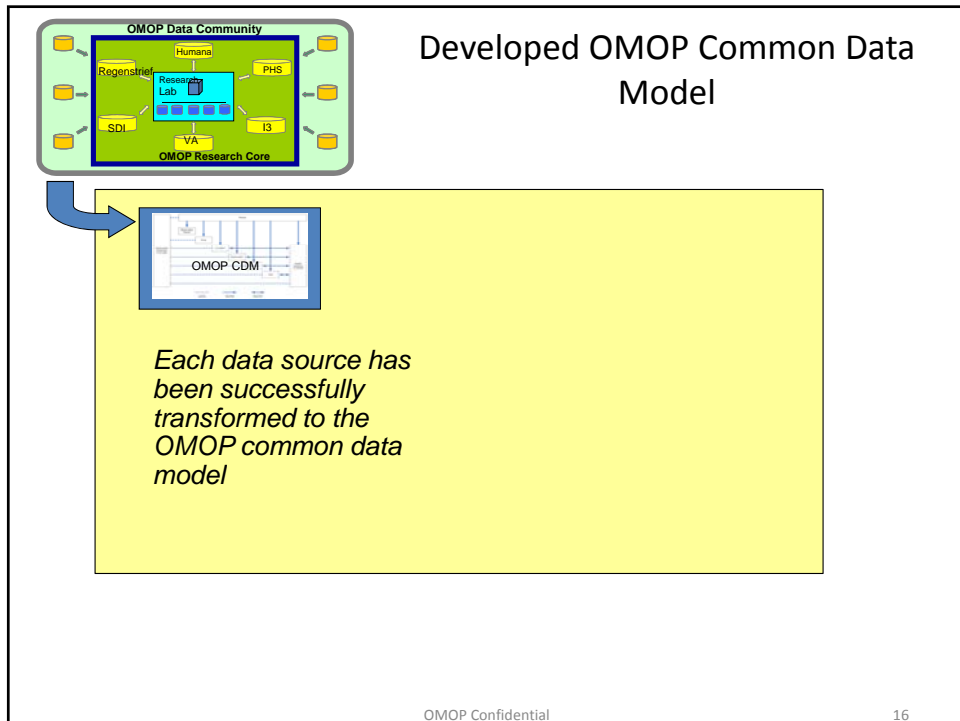
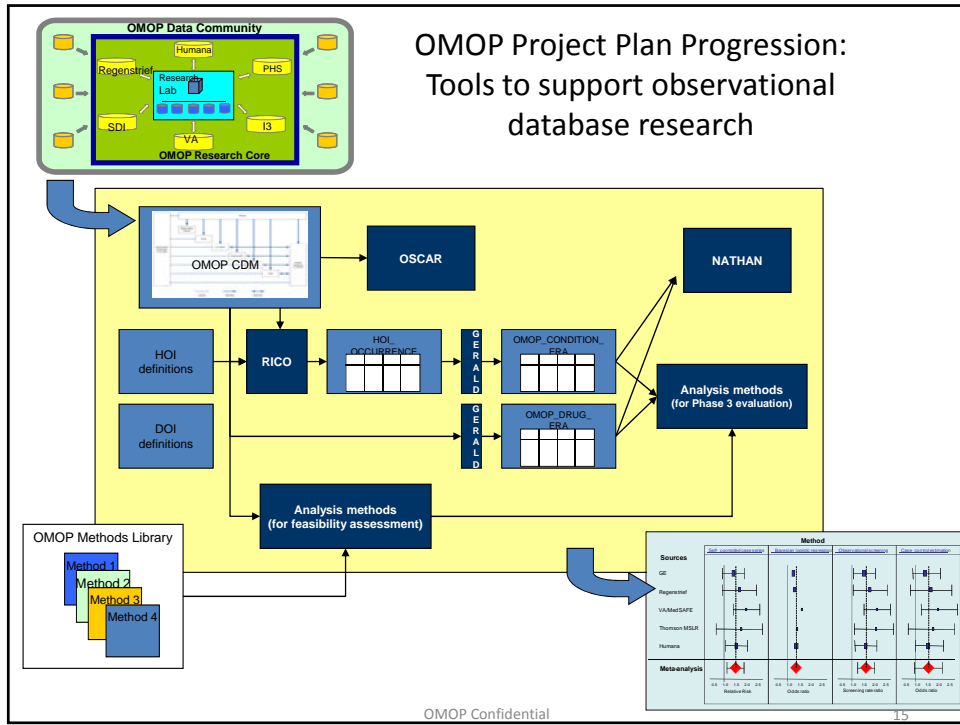
Analysis problems under study by OMOP

- **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 research experiment workflow





OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

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

The diagram illustrates the Common Data Model (CDM) structure. A central 'Person' entity is connected to several other entities: 'Observation Period', 'Drug', 'Condition', 'Observation', 'Procedure', and 'Visit'. A 'Terminology Dictionary (Concept)' is linked to 'Condition', 'Observation', 'Procedure', and 'Visit'. A 'Health Outcome of Interest' is also connected to 'Observation', 'Procedure', and 'Visit'. Relationships are defined as follows:

- Person to Observation Period: One-to-Many Relationship
- Person to Drug: One-to-Many Relationship
- Person to Condition: One-to-Many Relationship
- Person to Observation: One-to-Many Relationship
- Person to Procedure: One-to-Many Relationship
- Person to Visit: One-to-Many Relationship
- Condition to Observation: One-to-Many Relationship
- Observation to Procedure: One-to-Many Relationship
- Procedure to Visit: One-to-Many Relationship
- Observation to Health Outcome of Interest: One-to-Many Relationship
- Procedure to Health Outcome of Interest: One-to-Many Relationship
- Visit to Health Outcome of Interest: One-to-Many Relationship

 A legend at the bottom indicates:

- Dashed line: Concept Mapping
- Solid arrow: One-to-Many Relationship
- Double-headed arrow: Many-to-Many Relationship

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

OMOP Confidential 17

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Standardizing terminologies to accommodate disparate observational data sources

Standardizing conditions:

The diagram shows the standardization of conditions. Source codes include MedDRA, SNOMED-CT, ICD-9-CM, Read, and Oxmis. MedDRA is mapped to levels 1 through 5 (Low-level Terms, Preferred Terms, High Level Terms, High Level Group Terms, System Organ Class). SNOMED-CT is mapped to levels 1 through 3 (Low-level concepts, Higher-level classifications, Top-level classification). ICD-9-CM, Read, and Oxmis are mapped to level 1. A legend indicates:

- Solid arrow: Existing Mapping
- Dashed arrow: De Novo Mapping
- Double-headed arrow: Derived Mapping

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

Standardizing drugs:

The diagram shows the standardization of drugs. Source codes include GPI, NDC, Multum, HCPCS*, CPT-4*, and ICD-9-Proc*. These are mapped to RxNorm (Level 1 and 2), NDF-RT (Level 3 and 4), and RxNorm (Level 2). A legend indicates:

- Solid arrow: Existing Mapping
- Dashed arrow: De Novo Mapping
- Double-headed arrow: Derived Mapping

OMOP Confidential 18

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>

OMOP Confidential

19

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 Confidential

20

Developed Methods Library

OMOP Methods Library - Download Methods | Observational Medical Outcomes Partnership

OMOP Methods Library - Download Methods

OMOP is building a library of methods, developed for the OMOP Common Data Model, to address the analysis problems of Monitoring of Health Outcomes of Interest and Identification of Non-Specified Conditions. These methods will be tested across the OMOP Data Community. These methods are available under the [Apache public license](#).

If you would like to contribute to the methods, please contact OMOP by adding a new comment below.

Downloads Available

Guidelines

- **OMOP Methods development guidelines**
- **Disproportionality Analysis Method - OMOP Research Team**
- **Disproportionality Analysis Method specification** 7Dec 2009
- **Disproportionality Analysis Method Source Code and Examples** 27Feb 2010
- **Disproportionality Analysis Feasibility Test #1** 17Jan 2010
- **Disproportionality Analysis Feasibility Test #2** 17Jan 2010
- **Multi-Set Case-Control Estimation - OMOP Research Team**
- **Multi-set case-control Method specification** 7Dec 2009
- **Multi-set case-control Method Source Code and Examples** 2Feb 2010
- **Multi-set case-control Feasibility Test #1** 17Jan 2010
- **Multi-set case-control Feasibility Test #2** 17Jan 2010
- **Bayesian Logistic Regression - OMOP Research Team**
- **Bayesian logistic regression specification** 27Feb 2010

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

OMOP Confidential 21

Conduct Analyses to Evaluate Methods

OMOP Data Community

OMOP Research Core

OMOP CDM

OSCAR

NATHAN

HOI definitions

DOI definitions

RICO

HOI OCCURRENCE

GENERAL

OMOP_CONDITION_EPA

Analysis methods (for Phase 3 evaluation)

Analysis methods (for feasibility assessment)

OMOP Methods Library

Sources	Method			
	Self-controlled case series	Bayesian logistic regression	Observational scenario	Case-control estimation
GE				
Regenstrief				
VA/MedSAFE				
Thomson MSLR				
Humana				
Meta-analysis	Relative Risk	Odds ratio	Screening rate ratio	Odds ratio

OMOP Confidential 22

OBSERVATIONAL
MEDICAL
OUTCOMES
PARTNERSHIP

For further information

<http://omop.fnih.org>

OMOP Confidential

23