Advancing the Science for Active Surveillance: Rationale and Design for the Observational Medical Outcomes Partnership

Paul E. Stang, PhD; Patrick B. Ryan, MEng; Judith A. Racoosin, MD, MPH; J. Marc Overhage, MD, PhD; Abraham G. Hartzema, PharmD, MSPH, PhD; Christian Reich, MD, PhD; Emily Welebob, RN, MS; Thomas Scarnecchia, MS; and Janet Woodcock, MD

The U.S. Food and Drug Administration (FDA) Amendments Act of 2007 mandated that the FDA develop a system for using automated health care data to identify risks of marketed drugs and other medical products. The Observational Medical Outcomes Partnership is a public–private partnership among the FDA, academia, data owners, and the pharmaceutical industry that is responding to the need to advance the science of active medical product safety surveillance by using existing observational databases. The Observational Medical Outcomes Partnership's transparent, open innovation approach is designed to systematically and empirically study critical governance, data resource, and methodological issues and their interrelationships in establishing a viable national program of active drug safety surveillance by using observational data. This article describes the governance structure, data-access model, methods-testing approach, and technology development of this effort, as well as the work that has been initiated.

Ann Intern Med. 2010;153:600-606. www.annals.org
For author affiliations, see end of text.

When a new drug is approved, understanding of the product's safety profile is limited by the relatively small and narrowly defined study populations in the clinical trials required for approval. Uncommon adverse events are difficult to detect during premarket testing; therefore, developing methods to rapidly detect such events in the postmarket period is an urgent goal of the public health system. Currently, the U.S. Food and Drug Administration (FDA) relies primarily on the submission of spontaneous reports, which are often incomplete, reflect only a small percentage of actual events, and have limited use for outcomes with high background rates (1). Observational databases, containing administrative claims and electronic health records (EHRs), have frequently been used to characterize utilization patterns, track patient outcomes, and conduct formal pharmacoepidemiologic evaluation studies. However, the potential of these observational databases for active surveillance of medical products has not been substantively explored (2), except for vaccines (3). This gap in the understanding of how best to develop and apply active surveillance methods to these databases prompted the Observational Outcomes Medical Partnership (OMOP) project.

An active surveillance system involves a systematic process for analyzing multiple observational health care data sources to better understand the effects of medical products. In the FDA Amendments Act of 2007 (4), Congress mandated that the FDA collaborate with public, academic, and private entities to access disparate data sources and to validate ways to link and analyze safety data from multiple sources for medical product safety surveillance. An active surveillance system could potentially characterize known side effects, monitor preventable adverse events, and enhance the understanding of safety concerns emerging in the postmarket period by supplementing other sources of safety information (preclinical data, clinical trials, and spontaneous adverse event reporting).

The OMOP (http://omop.fnih.org), a public–private partnership among the FDA, academia, data owners, and the pharmaceutical industry and administered by the Foundation for the National Institutes of Health, was initiated to identify the needs of an active drug safety surveillance system and propose and test scientific methods and data infrastructure to address those needs. The OMOP research program consists of systematic and empirical investigations of the critical methodological and data resource issues within a specific technology architecture and governance model that is probably needed to establish a national medical product safety surveillance system. The ultimate goal of OMOP is to develop the necessary technology and methods to refine the secondary use of observational data for maximizing the benefit and minimizing the risk of pharmaceuticals.

See also:

Web-Only
Appendix
Appendix Table
Conversion of graphics into slides

Objectives and Design of OMOP

The OMOP was established to study the governance, data access, technology, and methods necessary to use existing observational databases for active drug safety and benefit monitoring. This work is being implemented over 2 years with the following goals: 1) define and test a pool of analytic methods that can be used to explore the relationships between drugs and health-related conditions across multiple types of observational data (administrative claims, inpatient and outpatient EHRs); 2) develop and test methods to apply to a network of central and distrib-

uted data sources for drug safety and effectiveness questions; 3) assess the performance of the analytic methods for 2 analysis problems—monitoring a defined set of risks and benefits that are considered to be "known" associations and identifying associations between drugs and outcomes that were not previously suspected; and 4) on the basis of the results of these analyses, determine how the results can shape the implementation of an active drug surveillance program.

Governance, Structure, and Oversight

The OMOP is a public–private partnership (5) that is an initiative of the Foundation for the National Institutes of Health, a 501(c)(3) organization. The primary funding comes from 17 corporate and nonprofit organizations in the pharmaceutical industry (**Appendix Table**, available at www.annals.org). These participants, in addition to other stakeholder groups, contribute intellectual property, expertise, and other in-kind resources.

The structure includes a central research core of scientists responsible for oversight of the OMOP program, creating and implementing the research protocols, and developing program code for the methods; a research laboratory that provides access to the 5 central databases; 6 funded research partners who represent a distributed network of diverse data sources, types, and populations; and funded methods collaborators.

The governance structure is designed to ensure that OMOP adheres to its guiding principles of transparency (work products are released in the public domain), collaboration, and empirical evaluation. The 10-member executive board is chaired by the director of the FDA's Center for Drug Evaluation and Research, with membership drawn from academia, regulatory agencies, the pharmaceutical industry, data holders, patient advocacy groups, and health care providers (**Appendix**, available at www.annals.org). Executive board members provide guidance on decisions regarding ethical and scientific concerns of the project and review and approve research plans, contracts, partnerships, and public announcements. They receive no compensation. The FDA representatives can veto executive board decisions if they believe the decisions are not in the best interest of the public. A 12-member scientific advisory board of methods experts and a 9-member health informatics advisory board provide independent expert review into the development of technology, data, and scientific methods (**Appendix**).

Data Access and Technology

Observational research requires substantial analytic capabilities to support very large databases and data-transformation activities. The OMOP research laboratory provides the core information technology services needed to support the OMOP research program and functions as a secure, interactive, centralized development center for methods research.

Data-Access Models

The OMOP is evaluating 2 data-access models that have been proposed for an active surveillance system: a distributed network and a centralized database. Its distributed network of 6 data holders represents a spectrum of experience, technical environments, and data types (EHR and administrative claims), as well as populations covered (**Table 1**). Each retains person-level data on site and provides only aggregate analyses to the research laboratory. With the support of the OMOP central team, each site is responsible for local conversion of its data to the common data model (CDM); implementation of methods; reporting of results; and collaboration in the development of methods, including feedback on local adaptations required. The distributed design is sensitive to the scientific, privacy, and ownership concerns of housing all data in a centralized data warehouse.

The centralized model is composed of 5 deidentified observational data sources (4 administrative claims and 1 EHR) that have been licensed and housed securely within the research laboratory and are made directly available to the research team (**Table 1**).

Simulated Data

A simulated data set enables benchmarking of methods against a database with known properties so that each method's performance characteristics (sensitivity, specificity, and predictive value) can be determined against a known "truth." The OMOP developed a way to generate simulated data sets (accessible from the OMOP Web site: http://omop.fnih.org) of user-defined size, data characteristics, and confounding to ensure that they are similar to actual data sets.

Transforming Automated Health Care Data to a Common Format

The OMOP has developed a CDM that does not alter the content of the data and allows researchers to develop analytic methods that can be run on any data source that adopts its format. The CDM was designed with broad stakeholder input to accommodate the key data elements expected to be necessary for active surveillance, inclusive of both EHR and administrative claims data. Participating organizations transform their data from their native formats into the specified structure and vocabulary that incorporates all relevant coding dictionaries (for example, International Classification of Diseases, Ninth Edition; Current Procedural Terminology, fourth edition; Systematized Nomenclature of Medicine; Logical Observation Identifiers Names and Codes) into a standardized terminology. This approach ensures that shared information, including methods, programs, benchmark tests, and results, can be consistently applied and interpreted across data sources. Many software tools (**Table 2**) provide a measure of quality as-

Table 1. **OMOP Data Community***

OMOP Collaborator (Representative Reference)	Type of Data	Geographic Coverage	Insurance Type	Available Lives, *n (millions)*
Central databases: commercially licensed databases				
GE Healthcare: Clinical Data Services (6)	EHR (primarily outpatient)	National	All payer types	11.2
Thomson Reuters				
Commercial claims (7)	Administrative claims	National	Multiple private insurers	58
Medicare supplement (7)	Administrative claims	National	Medicare supplement	4.4
Medicaid (7)	Administrative claims	National (multiple-states Medicaid)	Medicaid	11.1
MarketScan Lab (7)	Subpopulation with administrative claims and laboratory results	National	Multiple	1.5
Total				86
Distributed partners: database collaborating sites				
Indiana University–Regenstrief Institute (8)	Integrated health care exchange: administrative claims and EHR	Indianapolis metropolitan area	All	9.4
i3 Drug Safety–Ingenix NHI (9)	Administrative claims	National	Single private insurer	50
Partners Healthcare Systems (10)	EHR	Boston regional area	Single provider organization	5.0
University of Miami–Humana HSRC (11)	Administrative claims, including Medicare advantage and prescription plans	National	Single private insurer	6.5
SDI Health (12)	Administrative claims from point of care with EHR in subset	National	All payer types	160
Total				231
Federal collaborators				
U.S. Department of VA (13)	EHR at VA facilities	National	Single payer/delivery system	7.8
Total				7.8
Grand total	–	–	–	325

EHR = electronic health record; HSRC = Health Services Research Center; NHI = Normative Health Information; OMOP = Observational Medical Outcomes Partnership; VA = Veterans Affairs.
* Including reference describing database or representative research using database.

surance for the CDM transformation process and critical background characteristics of the populations in the databases.

Health Outcome of Interest Definitions

A critical feature of the OMOP studies is the characterization of what are thought to be known associations or drug–event pairs listed in a product package insert that have been confirmed in observational database studies. The particular health outcomes of interest for OMOP were selected on the basis of their representation of the spectrum of adverse events (for example, with regard to background rate, time to onset, or presence in a boxed warning on the product labeling) or their likelihood of being the focus of ongoing drug safety surveillance (**Table 3**) (Stang PE, Ryan PB, Dusetzina S, et al. Health outcomes of interest in ob-

Table 2. **OMOP Tools Under Development and Their Application**

Name	Application
OSIM	Open-source software application, written in R, that allows users to create simulated data sets that conform to the OMOP CDM. The simulation creates hypothetical persons with fictitious drug exposure and conditions, with known characteristics that represent the types of scenarios expected in real observational sources.
OSCAR	An SAS program that creates descriptive statistics, allowing the following functions: summarizing available data from a given source within the OMOP CDM; providing context for interpreting and analyzing findings of drug safety studies; facilitating comparisons between data sources; enabling comparison of overall database to specific subpopulations of interest; and supporting validation of transformation from raw data to OMOP CDM.
NATHAN	An SAS program extension of OSCAR, creating a standardized summary providing some context and expected rates of drug utilization and condition occurrence to facilitate the interpretation of benefit and risk information and generate a standardized report summarizing characteristics about the population of interest, including demographic factors (age and sex); comorbid conditions and concomitant medications; and health service utilization before, during, and after the event onset.
RICO	A procedure that standardizes patient cohort selection. Standard "cohort definitions" are created by using criteria specified in input parameters. These cohort definitions are input into RICO, and patients meeting the criteria can be automatically and rapidly selected from any database conforming to the OMOP CDM.

CDM = common data model; NATHAN = Natural History Analysis; OMOP = Observational Medical Outcomes Partnership; OSCAR = Observational Source Characteristics Analysis Report; OSIM = Observational Medical Dataset Simulator; RICO = Regularized Identification of Cohorts.

Table 3. **Drug Condition Pairs, by Outcome Type and Rate: The OMOP Test Cases**

Outcome Type	Outcome Rate in Population	
	Rare	Common
Safety	Angioedema: ACE inhibitors Renal failure: amphotericin B Acute liver injury: antibiotics (erythromycin, sulfonamides, tetracyclines) Aplastic anemia: antiepileptics (carbamazepine, phenytoin)	Hip fracture: benzodiazepines GI ulcer hospitalizations: alendronate MI: tricyclic antidepressants, typical antipsychotics Hemorrhage: warfarin
Benefit	–	Reduced hospitalizations: ACE inhibitors Lower mortality after MI: β-blockers

ACE = angiotensin-converting enzyme; GI = gastrointestinal; MI = myocardial infarction; OMOP = Observational Medical Outcomes Partnership.

servational data: issues in identifying definitions in the literature. Unpublished data.). The 2 benefit case examples were drawn from a similar process. Multiple definitions were derived from systematic reviews of observational research, clinical diagnosis guidelines, and database characteristics; the effect on method performance across variations in the definitions will be tested.

Analyses and Methods

Two distinct types of analyses in the OMOP's research program constitute the general surveillance scenarios in observational data: 1) identification of known or suspected associations and 2) identification of drug–outcome associations that were not previously suspected. Both surveillance scenarios will be evaluated across the multiple data sources. Each type of analysis presents different challenges, requires different algorithms, and uses different data elements. In the first analysis, the surveillance is focused on health outcomes of interest that are identified during the clinical development program or the postmarket period, are known to be associated with the class of compounds, are biologically or theoretically of interest, or represent a toxicity generally associated with medications (such as acute hepatic injury). A desirable characteristic of a method is its ability to discriminate between a known association and a negative control (that is, a drug–outcome pair for which no evidence suggests an association based on the product labeling and the medical literature).

The second analytic scenario is the identification of drug–outcome associations that were not previously suspected (nonspecified associations). The OMOP will assess the accuracy of each method in identifying nonspecified associations by comparing the set of associations identified by the method against the gold standard of the adverse events listed in a most recently approved product labeling of the drug. We recognize that a drug's product labeling may not reflect all of the possible adverse events that have been observed, but it does represent FDA's most comprehensive description of a drug's adverse effect profile.

Methods for Inclusion in OMOP

A list of potential methods for active surveillance was established on the basis of the published literature, through solicitation from the methods community, and through a methods competition (the "OMOP Cup"). More than 10 categories of methods emerged and are included in the OMOP research program, developed by a broad community of methodologists. Many of these methods have not been applied to systematic surveillance of health care data (**Table 4**). They are being catalogued in a publicly accessible library that includes the software code.

OMOP and the Process of Safety Surveillance

The program of research for the OMOP is intended to clarify many issues in our understanding of the methods and data needed for active drug safety surveillance, including:

Validation: Can we confirm that the native data transformed appropriately into the CDM standardized format? Can we confirm that vocabulary mapping occurs in a consistent manner?

Feasibility: Was each method able to run against each data set and in each computing environment?

Performance: How well did each method work, and did it return results consistent with expectations?

Ultimately, these findings must be integrated into the other research being conducted in the active medical product surveillance field, because the OMOP will probably not provide definitive answers. Under the FDA's Sentinel Initiative, a distributed system will be designed to implement the congressional mandate that FDA create an electronic system to analyze safety data from multiple sources for medical product safety surveillance. In concert with that initiative, OMOP is assessing data needs and objectively measure method performance to help decision makers determine the requirements of an active medical product surveillance system.

We can best learn how to identify new drug–outcome associations that were not previously suspected by refining methods for active surveillance on well-established, known drug–health outcome of interest associations. Focusing on known associations maximizes our understanding of the

Table 4. **Inventory of Analysis Procedures***

Program (Reference)	Collaborator	Description
Disproportionality analyses		
Disproportionality analysis (14–16)	Columbia University	Methods adapted from data mining of spontaneous adverse event reports, where drug–condition pairs are identified if they co-occur disproportionately more frequently than expected if the drug and condition were independent. Metrics include the MGPS, PRR, ROR, and BCPNN.
Disproportionality analysis; Bayesian method (17)	Merck & Co.	An alternative Bayesian approach to disproportionality analysis to control the false-positive and false-negative rates.
Temporal pattern discovery (18, 19)	Uppsala Monitoring Centre	This is a novel method for event history data, focusing explicitly on the detailed temporal relationship between pairs of events. The proposed measure contrasts the observed-to-expected ratio in a period of interest with that in a predefined control period.
Case-based approaches		
Multiset case–control estimation (20)	Columbia University–GlaxoSmithKline	The program leverages the basic design of a case–control study to enable estimates of drug–condition associations across a large set of drugs and conditions. The algorithm can estimate an odds ratio simultaneously for multiple conditions and allows all exposures to be evaluated for each outcome.
Case–control surveillance (21)	Eli Lilly and Company	The program applies a case–control surveillance design to estimate odds ratios for drug–condition effects, where cases are matched to controls by age, sex, location, and race.
Self-controlled case series (22)	Columbia University	The method estimates the association between a transient exposure and adverse event using only cases; no separate controls are required because each case acts as its own control.
Case-crossover (23)	University of Utah	The design uses within-participant comparisons of drug exposures over time to estimate the rate ratio of the outcome associated with the drug under study.
Exposure-based approaches		
Observational screening (24, 25)	ProSanos	This is an extension of a traditional cohort epidemiology design where the rate of ADEs can be compared across groups of patients exposed to different medications, allowing comparisons within a cohort population, between treatments, as well as relative to the overall population at large.
High-throughput screening (26)	Regenstrief Institute–Indiana University School of Medicine	This method calculates relative risk and incidence rate differences between exposure cohorts relative to population estimates.
High-dimensional propensity scoring (27)	University of North Carolina at Chapel Hill–SAS Institute	This is a multistep algorithm to implement high-dimensional proxy adjustment in observational data. Used in conjunction with a new-user cohort design, it offers a novel approach to minimizing confounding when assessing the relative association between patients exposed to alternative medications and the occurrence of a health outcome of interest.
Local control (28)	Risk-Benefit Statistics	Local control is a robust alternative to traditional covariate adjustment methods using multivariable statistical models by making treatment comparisons only within clusters of relatively well-matched patients.
Sequential methods		
MaxSPRT (29, 30)	Harvard Pilgrim Health Care–Group Health Cooperative	MaxSPRT is a sequential analysis method designed for continuous or frequent (e.g., weekly) monitoring of a potential elevated risk for an adverse event after introduction of a drug or vaccine of interest.
CSSP (31)	Harvard Pilgrim Health Care	CSSP is a practical group sequential method with a finite number of interim tests to determine whether the drug of interest leads to an elevated risk compared with a comparator drug. It is designed for settings in which information for both the drug of interest and the comparator drug accumulates over time.
Other methods		
Bayesian logistic regression (32, 33)	Columbia University	This is a high-dimensional statistical method that is scalable to a substantial number of covariates, accommodating all drugs and conditions in a single model to predict occurrence of ADEs. The Bayesian approach to logistic regression has several advantages, including avoidance of overfitting, efficiency during model prediction time, and scalability to large numbers of covariates (see also www.bayesianregression.org).
Statistical relational learning (34, 35)	University of Wisconsin	This method adapts a machine-learning approach to work directly with relational data distributed across many tables to extract "rules" that define observed phenomena, such as drug–condition relationships.

ADE = adverse drug event; BCPNN = Bayesian confidence propagation neural network; CSSP = conditional sequential sampling procedure; MaxSPRT = maximized sequential probability ratio test; MGPS = Multi-item Gamma-Poisson Shrinker; PRR = proportional reporting ratios; ROR = reporting odds ratio.
* Detailed white papers can be found at http://omop.fnih.org/MethodsLibrary for each program. For some programs, the references provided are representative of the published application of the method, which may have required adaptation to be applicable for active surveillance.

methods' performance. However, these associations may not currently be as robust as demonstrated in earlier observational studies because practice patterns may have adjusted to minimize risk for the adverse outcomes or the populations exposed to the drugs of interest may have changed.

Our work will explore the relative value of the clinical information contained in the EHR for active surveillance purposes compared with that available in administrative claims databases. Holbrook and colleagues (36) found that EHRs contain most of the data fields for optimal routine pharmacosurveillance and the potential richness of the

EHR may help overcome some of the coding issues in claims databases (coding for reimbursement vs. reflecting actual clinical care). This work may also help inform the goals of "meaningful use" (37) to "improve population and public health." The OMOP's use of a common data structure, inclusive of any type of data (EHR, claims), facilitates the development and refinement of methods for active surveillance, which will contribute to the improvement of public health.

Although we are assessing the computational feasibility of the methods, this is purely a technical test of feasibility, and a method may fail and be eliminated from further consideration despite having desirable performance characteristics. Future research should reconsider our complete pool of methods.

Attention to the assessment of benefit is particularly relevant given the increasing interest in using these data sources to address questions of comparative effectiveness. The OMOP program evaluates benefits in a limited manner. Developing a systematic way to analyze benefit in observational data is appealing, but the broader spectrum of benefits (quality of life, productivity, and functioning) are poorly represented even in EHRs. In some cases, benefit can be defined by the relative absence of "risk." This paradigm would allow implementation of the OMOP process for comparative effectiveness analyses in cases where the presence or absence of a clinical event represents the relevant end point for analysis (for example, the relative risk for bleeding with warfarin compared with a newly developed anticoagulant).

Other efforts (for example, Exploring and Understanding Adverse Drug Reactions by Integrative Mining of Clinical Records and Biomedical Knowledge Project [EU-ADR] linking EHRs from 4 countries, and FDA mini-Sentinel pilot, a funded coordinating center for a consortium of automated health care databases) are also developing data networks and methods to conduct active surveillance and will provide an opportunity to test and refine the methods, data, or infrastructure developed by OMOP. Our hope is that OMOP will stimulate the establishment of a vibrant observational science (or epidemiology–informatics) research community, supported in part by our efforts to make the research laboratory available to qualified methods developers. In the end, our efforts will pay off handsomely if we can engender interest in and support for sustaining similar research initiatives and create more opportunities for education and training of scientists, decision makers, and the public in using and understanding observational data for active safety surveillance.

From Johnson & Johnson Pharmaceutical Research and Development, Titusville, New Jersey; Observational Medical Outcomes Partnership, Foundation for the National Institutes of Health, Bethesda, Maryland; GlaxoSmithKline, Research Triangle Park, North Carolina; Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina; Center for Drug Evaluation and Research, U.S. Food and Drug Administration, Silver Spring, Maryland; Regenstrief Institute and Indiana University School of Medicine, Indianapolis, Indiana; College of Pharmacy, University of Florida, Gainesville, Florida; and Digital Aurora, Manchester, Vermont.

Disclaimer: This article expresses the views of the authors and does not necessarily represent those of their affiliated organizations.

Acknowledgment: The OMOP thanks Jeffrey S. Dubin, MD, MBA, for his clinical and editorial input.

Grant Support: The OMOP is funded by the Foundation for the National Institutes of Health, with contributions from a consortium of 17 organizations (**Appendix Table**, available at www.annals.org). Three members of the research team (Dr. Stang, Mr. Ryan, and Dr. Racoosin) serve in-kind and are not compensated for their participation.

Potential Conflicts of Interest: Disclosures can be viewed at www.acponline.org/authors/icmje/ConflictOfInterestForms.do?msNum=M10-1147.

Requests for Single Reprints: Paul E. Stang, PhD, Johnson & Johnson, 1125 Trenton-Harbourton Road, PO Box 200, MS K304, Titusville, NJ 08560; e-mail, PStang@its.jnj.com.

Current author addresses and author contributions are available at www.annals.org.

References

1. **Bennett CL, Nebeker JR, Yarnold PR, Tigue CC, Dorr DA, McKoy JM, et al.** Evaluation of serious adverse drug reactions: a proactive pharmacovigilance program (RADAR) vs safety activities conducted by the Food and Drug Administration and pharmaceutical manufacturers. Arch Intern Med. 2007;167:1041-9. [PMID: 17533207]

2. **Institute of Medicine.** The Future of Drug Safety: Promoting and Protecting the Health of the Public. Washington, DC: National Academies Pr; 2006.

3. **Lieu TA, Kulldorff M, Davis RL, Lewis EM, Weintraub E, Yih K, et al; for the Vaccine Safety Datalink Rapid Cycle Analysis Team.** Real-time vaccine safety surveillance for the early detection of adverse events. Med Care. 2007;45: S89-95. [PMID: 17909389]

4. Food and Drug Administration Amendments Act of 2007, Pub. L. No. 110-85, § 905.

5. **National Council of Public Private Partnerships.** How PPPs work. Accessed at www.ncppp.org/howpart/index.shtml on 30 January 2010.

6. **Gill JM, Chen YX, Lieberman MI.** Management of depression in ambulatory care for patients with medical co-morbidities: a study from a national Electronic Health Record (EHR) network. Int J Psychiatry Med. 2008;38:203-15. [PMID: 18724571]

7. **Adamson DM, Chang S, Hansen LG.** Health Research Data for the Real World: The MarketScan Databases. White Paper. New York: Thomson Healthcare; January 2008. Accessed at http://home.thomsonhealthcare.com/uploadedFiles/docs/2008HealthResearchDatafortheRealWorldThe%20MarketScanDatabases(1).pdf on 15 August 2010.

8. **McDonald CJ, Overhage JM, Barnes M, Schadow G, Blevins L, Dexter PR, et al; INPC Management Committee.** The Indiana network for patient care: a working local health information infrastructure. An example of a working infrastructure collaboration that links data from five health systems and hundreds of millions of entries. Health Aff (Millwood). 2005;24:1214-20. [PMID: 16162565]

9. **Shatin D, Rawson NS, Stergachis A.** UnitedHealth Group. In: Strom B, ed. Pharmacoepidemiology. 4th ed. West Sussex, UK: J Wiley; 2005:271-80.

10. **Brownstein JS, Murphy SN, Goldfine AB, Grant RW, Sordo M, Gainer V, et al.** Rapid identification of myocardial infarction risk associated with diabetes medications using electronic medical records. Diabetes Care. 2010;33:526-31. [PMID: 20009093]

11. **Palacio A, Uribe CL, Li H, Hanna J, Deminski M, Alvir J, et al.** Financial and clinical characteristics of fibromyalgia: a case-control comparison. Am J Manag Care. 2010;16:S118-25. [PMID: 20586520]
12. **Pamer CA, Hammad TA, Wu YT, Kaplan S, Rochester G, Governale L, et al.** Changes in US antidepressant and antipsychotic prescription patterns during a period of FDA actions. Pharmacoepidemiol Drug Saf. 2010;19:158-74. [PMID: 20049836]
13. **Berlowitz DR, Miller DR, Oliveria SA, Cunningham F, Gomez-Caminero A, Rothendler JA.** Differential associations of beta-blockers with hemorrhagic events for chronic heart failure patients on warfarin. Pharmacoepidemiol Drug Saf. 2006;15:799-807. [PMID: 16892457]
14. **Curtis JR, Cheng H, Delzell E, Fram D, Kilgore M, Saag K, et al.** Adaptation of Bayesian data mining algorithms to longitudinal claims data: coxib safety as an example. Med Care. 2008;46:969-75. [PMID: 18725852]
15. **Hochberg AM, Hauben M, Pearson RK, O'Hara DJ, Reisinger SJ, Goldsmith DI, et al.** An evaluation of three signal-detection algorithms using a highly inclusive reference event database. Drug Saf. 2009;32:509-25. [PMID: 19459718]
16. **Almenoff JS, DuMouchel W, Kindman LA, Yang X, Fram D.** Disproportionality analysis using empirical Bayes data mining: a tool for the evaluation of drug interactions in the post-marketing setting. Pharmacoepidemiol Drug Saf. 2003;12:517-21. [PMID: 14513665]
17. **Gould AL.** Accounting for multiplicity in the evaluation of "signals" obtained by data mining from spontaneous report adverse event databases. Biom J. 2007; 49:151-65. [PMID: 17342957]
18. **Norén GN, Bate A, Hopstadius J, Star K, Edwards IR.** Temporal pattern discovery for trends and transient effects: its application to patient records. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Las Vegas, Nevada, 24–27 August 2008. New York: Association for Computing Machinery; 2008.
19. **Norén GN, Hopstadius J, Bate A, Star K, Edwards I.** Temporal pattern discovery in longitudinal electronic patient records. Data Min Knowl Discov. 2010;20: 361-87.
20. **Breslow NE, Day NE.** Statistical Methods in Cancer Research. Volume 1: The Analysis of Case-Control Studies. Lyon, France: International Agency for Research on Cancer; 1993.
21. **Rosenberg L, Coogan PF, Palmer JR.** Case-control surveillance. In: Strom B, ed. Pharmacoepidemiology. 4th ed. West Sussex, UK: J Wiley; 2005:185-202.
22. **Whitaker HJ, Hocine MN, Farrington CP.** The methodology of self-controlled case series studies. Stat Methods Med Res. 2009;18:7-26. [PMID: 18562396]
23. **Schneeweiss S, Stürmer T, Maclure M.** Case-crossover and case-time-control designs as alternatives in pharmacoepidemiologic research. Pharmacoepidemiol Drug Saf. 1997;6 Suppl 3:S51-9. [PMID: 15073755]
24. **Ryan PB, Powell GE, Pattishall EN, Beach KJ.** Performance of screening multiple observational databases for active drug safety surveillance. Poster Presented at the 25 Annual Meeting of the International Society of Pharmacoepidemiology, Providence, Rhode Island, 16–19 August 2009. Accessed at www.prosanos.com/uploaded/Ryan_ISPE2009_handout_18aug2009.pdf on 19 August 2010.
25. **Ryan PB, Powell GE.** Exploring candidate differences between drug cohorts prior to exposure: a systematic approach using multiple observational databases. Poster Presented at the 13th International Society of Pharmacoeconomics and Outcomes Research Annual Meeting, Toronto, Ontario, Canada, 3–7 May 2008.
26. **Janzen WP, ed.** High Throughput Screening: Methods and Protocols. 2nd ed. Totowa, NJ: Humana Pr; 2002.
27. **Schneeweiss S, Rassen JA, Glynn RJ, Avorn J, Mogun H, Brookhart MA.** High-dimensional propensity score adjustment in studies of treatment effects using health care claims data. Epidemiology. 2009;20:512-22. [PMID: 19487948]
28. **Faries D, Leon A, Haro J, Obenchain RL.** The local control approach using JMP. In: Analysis of Observational Health Care Data Using SAS. Cary, NC: SAS Pr; 2010.
29. **Kulldorff M, Davis R, Kolczak M, Lewis E, Lieu T, Platt R.** A maximized sequential probability ratio test for drug and vaccine safety surveillance. Working Paper. Boston: Harvard Medical School Department of Ambulatory Care and Prevention Working Paper; 2007. Accessed at www.populationmedicine.org/content/docs/maxSPRTsubm.pdf on 9 September 2010.
30. **Brown JS, Kulldorff M, Chan KA, Davis RL, Graham D, Pettus PT, et al.** Early detection of adverse drug events within population-based health networks: application of sequential testing methods. Pharmacoepidemiol Drug Saf. 2007; 16:1275-84. [PMID: 17955500]
31. **Li L.** A conditional sequential sampling procedure for drug safety surveillance. Stat Med. 2009;28:3124-38. [PMID: 19691034]
32. **Genkin A, Lewis D, Madigan D.** Large-scale Bayesian logistic regression for text categorization. Technometrics. 2007;49:291-304.
33. **Hauben M, Madigan D, Gerrits CM, Walsh L, Van Puijenbroek EP.** The role of data mining in pharmacovigilance. Expert Opin Drug Saf. 2005;4:929-48. [PMID: 16111454]
34. **Davis J, Burnside E, Dutra I, Page D, Ramakrishnan R, Santos Costa V, et al.** View learning for statistical relational learning: with an application to mammography. In: Proceedings of the International Joint Conference on Artificial Intelligence, Edinburgh, United Kingdom, 30 July–5August 2005. Menlo Park, CA: International Joint Conference on Artificial Intelligence; 2005:677-83.
35. **Burnside ES, Davis J, Chhatwal J, Alagoz O, Lindstrom MJ, Geller BM, et al.** Probabilistic computer model developed from clinical data in national mammography database format to classify mammographic findings. Radiology. 2009;251:663-72. [PMID: 19366902]
36. **Holbrook A, Grootendorst P, Willison D, Goldsmith C, Sebaldt R, Keshavjee K.** Can current electronic systems meet drug safety and effectiveness requirements? AMIA Annu Symp Proc. 2005:335-9. [PMID: 16779057]
37. 42 C.F.R. § 412, § 413, § 422, §495 (2010).

Current Author Addresses: Dr. Stang: Johnson & Johnson, 1125 Trenton-Harbourton Road, PO Box 200, MS K304, Titusville, NJ 08560.
Mr. Ryan, Dr. Reich, and Ms. Welebob: Foundation for the National Institutes of Health, 9650 Rockville Pike, Bethesda, MD 20814.
Dr. Racoosin: U.S. Food and Drug Administration, 10903 New Hampshire Avenue, WO51-6346, Silver Spring, MD 20993-0002.
Dr. Overhage: Medical Informatics, Regenstrief Institute, Inc., Health Information and Translational Sciences Building (HITS), 410 West 10th Street, Suite 2000, Indianapolis, IN 46202.
Dr. Hartzema: College of Public Health and Health Professions, University of Florida, HPNP, Room 3339, 101 South Newell Drive, PO Box 100496, Gainesville, FL 32610-0496.
Mr. Scarnecchia: PO Box 686, Manchester, VT 05254.
Dr. Woodcock: U.S. Food and Drug Administration, 10903 New Hampshire Avenue, Building 51, Room 6133, Silver Spring, MD 20993.

Author Contributions: Conception and design: P.E. Stang, P.B. Ryan, J.M. Overhage, A.G. Hartzema, C. Reich.
Analysis and interpretation of the data: P.E. Stang, P.B. Ryan, J.A. Racoosin, J.M. Overhage, A.G. Hartzema, C. Reich, E. Welebob.
Drafting of the article: P.E. Stang, P.B. Ryan, J.M. Overhage, E. Welebob, T. Scarnecchia.
Critical revision of the article for important intellectual content: P.E. Stang, P.B. Ryan, J.A. Racoosin, J.M. Overhage, A.G. Hartzema, C. Reich, J. Woodcock.
Final approval of the article: P.E. Stang, P.B. Ryan, J.A. Racoosin, J.M. Overhage, A.G. Hartzema, T. Scarnecchia, J. Woodcock.
Provision of study materials or patients: P.B. Ryan.
Statistical expertise: P.E. Stang, P.B. Ryan.
Obtaining of funding: P.E. Stang, P.B. Ryan, T. Scarnecchia.
Administrative, technical, or logistic support: P.E. Stang, P.B. Ryan, C. Reich, E. Welebob, T. Scarnecchia, J. Woodcock.
Collection and assembly of data: P.E. Stang, P.B. Ryan, J.M. Overhage, C. Reich, E. Welebob.

Appendix

OMOP Executive Board Members

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

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

Sherine E. Gabriel, MD, MSc, William J. and Charles H. Mayo Professor, 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 (Acting) for Science and Public Health, U.S. 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 Pharmocogenomics, Children's Mercy Hospital; Dean Emeritus, Dartmouth Medical School

Brian L. 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

OMOP Scientific Advisory Board Members

Elizabeth B. Andrews, MPH, PhD, Vice President, Pharmacoepidemiology and Risk Management, RTI Health Solutions; Adjunct Associate Professor, University of North Carolina at Chapel Hill Schools of Public Health and Pharmacy

Andrew Bate, PhD, Senior Director, Quantitative Epidemiologist, Pfizer

Jesse Berlin, ScD, Vice President, Epidemiology, Johnson & Johnson Pharmaceutical Research and Development

Robert Lowell Davis, MD, MPH, Director of Research, Center for Health Research Southeast, Kaiser Permanente Georgia

Steven D. Findlay, Senior Health Policy Analyst, Consumers Union

Sean Hennessy, PharmD, PhD, Assistant Professor of Epidemiology and of Pharmacology, University of Pennsylvania School of Medicine

Michael S. Katz, Vice President; Chair, Patient Advisory Board; International Myeloma Foundation; Coalition of Cancer Cooperative Groups

Allen A. Mitchell, MD, Professor of Epidemiology & Pediatrics, Boston University Schools of Public Health & Medicine; Director, Slone Epidemiology Center at Boston University

David Page, PhD, Professor, Department of Biostatistics & Medical Informatics, University of Wisconsin–Madison

Kenneth J. Rothman, DrPH, RTI Health Solutions, RTI International; Distinguished Fellow and Vice President, Epidemiology Research, Professor of Epidemiology and Medicine, Boston University

Judy A. Staffa, PhD, RPh, Associate Director for Regulatory Research, Center for Drug Evaluation and Research Office of Surveillance and Epidemiology, U.S. Food and Drug Administration

Alexander M. Walker, MD, DrPH, Principal, World Health Information Science Consultants; Adjunct Professor, Department of Epidemiology Harvard School of Public Health

OMOP Health Care Informatics Advisory Board Members

COL Kevin Abbott, MD, MPH, Staff Nephrologist, Washington, DC

Jeffrey S. Brown, PhD, Lecturer, Department of Ambulatory Care and Prevention, Harvard Medical School; Harvard Pilgrim Health Care; Director, HMO Research Network Center for Education and Research on Therapeutics Data Coordinating

Center; Research Director, Therapeutics Research and Infectious Disease, Epidemiology

Stanley M. Huff, MD, Chief Medical Informatics Officer, Intermountain Healthcare; Professor, University of Utah School of Medicine

Diane T. MacKinnon, Patient Consultant, Center for Drug Evaluation and Research, U.S. Food and Drug Administration

Kenneth D. Mandl, MD, MPH, Associate Professor, Harvard Medical School; Director, Intelligent Health Laboratory, Children's Hospital Informatics Program, Children's Hospital Boston

Clement J. McDonald, MD, Director, Lister Hill Center for Biomedical Communications

David S. Memel, MD, MS, MBA, Chief Medical Officer, Senior Vice President of Analytics, LifeCare

Mitra Rocca, MSc, Associate Director, Medical Informatics, U.S. Food and Drug Administration

Robert Thwaites, Senior Executive Director, United BioSource Corporation–Europe

Appendix Table. **Funding Organizations and Stakeholder Groups Providing Resources to OMOP**

Organization	Activity	Funding Provided to FNIH	Provided In-Kind Contributions	Grant Support	Contract/ Compensated
Abbott	Funding	X			
Amgen	Funding	X			
AstraZeneca	Funding	X			
Bayer Healthcare Pharmaceuticals	Funding	X			
Bristol-Myers Squibb	Funding	X			
Columbia University	Principal investigator and programming and statistical analysis				X
Computer Sciences Corporation	Research laboratory				X
Department of Veterans Affairs PBM Center for Medication Safety	Distributed research partner			X	
Eli Lilly & Company	Funding and methods collaborator	X	X		
U.S. Food and Drug Administration	Principal investigator, advisory board, and executive board		X		
GE Healthcare	Research laboratory				X
GlaxoSmithKline	Funding and research investigator	X	X		
The GPRD Group of the Medicines and Healthcare products Regulatory Agency	Consulting				X
Harvard Pilgrim Health Care Institute	Methods collaborator, advisory board, and executive board		X	X	
University of Miami–Humana Health Services Research Center	Distributed research partner			X	
i3 Drug Safety	Distributed research partner			X	
Indiana University–Regenstrief Institute	Principal investigator, distributed research partner, and methods collaborator			X	X
Johnson & Johnson	Funding, principal investigator, and advisory board	X	X		
Lundbeck	Funding	X			
Merck & Co.	Funding, methods collaborator, and health outcomes of interest library	X	X		
Novartis Pharmaceuticals	Funding	X			
Partners HealthCare System	Distributed research partner			X	
Pfizer	Funding and advisory board	X		X	
Pharmaceutical Research Manufacturers of America	Funding and executive board	X			
ProSanos	Simulated data and methods collaborator			X	X
Risk-Benefit Statistics	Methods collaborator				X
Roche	Funding	X			
RTI International	Health outcomes of interest library and advisory board		X		X
sanofi-aventis	Funding	X			
Schering-Plough	Funding	X			
SDI Health	Distributed research partner			X	
Takeda	Funding	X			
Thomson Reuters	Research laboratory				X
United BioSource	Health outcomes of interest library and advisory board		X		X
University of Florida	Principal investigator				X
University of North Carolina at Chapel Hill and SAS Institute	Methods collaborator				X
University of Utah	Methods collaborator			X	
University of Wisconsin–Madison	Methods collaborator and advisory board			X	
The Uppsala Monitoring Centre	Methods collaborator			X	

FNIH = Foundation for the National Institutes of Health; GPRD = General Practice Research Database; OMOP = Observational Medical Outcomes Partnership; PBM = pharmacy benefits manager.