Volume 2, Issue 3 October 2010

OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

Partnership Update

Announcements

In this quarter's newsletter, we cover two topics: the status of the current research plan and plans for the future. The Observational Medical Outcomes Partnership (OMOP) is on track to successfully complete the goals outlined in the original research plan by the first quarter of 2011. We explore the future of OMOP and the Foundation for the National Institutes of Health's efforts to broaden the scope of funding to sustain OMOP beyond the original time frame.

We are pleased to announce the addition of two new members to our Executive Board. Dr. Marcus Wilson of HealthCore and Dr. Steven Jacobsen of Southern California Permanente Medical Group joined us in late September. We will introduce Dr. Wilson and Dr. Jacobsen in the December issue of the OMOP Newsletter.

OMOP will hold the second annual OMOP Symposium on January 11, 2011, in Washington, DC. Additionally, OMOP will host a series of workshops and tutorials in the near future. Please stay tuned to the OMOP website (omop.fnih.org) for the latest updates.

In This Issue

OMOP Status Update

Key Developments

The ultimate goal of the Observational Medical Outcomes Partnership (OMOP) is to inform the implementation of a systematic surveillance system with empirical evidence of the appropriate methods, data, and infrastructure necessary to identify and track potential relevant safety issues (and, to a lesser extent, drug benefit). OMOP is empirically studying the contribution of and synergies among data, technology, methods, and governance required to establish an active surveillance system for pharmaceutical safety using existing observational databases. In addition, OMOP is also evaluating the use of these data resources to better understand its utility in assessing benefits. This work has been conducted over the past two years with the following research goals:

- Defining and testing a pool of analytical methods that can be used to explore the relationships among drugs and health-related conditions across multiple types of observational data (administrative claims and inpatient and outpatient electronic health records)
- Developing a set of tools for enabling a network of disparate data sources to test methods and evaluate alternative data access models, including distributed network and centralized systems
- Applying these tools and performing analyses against a limited set of specific observational data sources that (a) provide an immediate, practical test of the research objectives and (b) represent the major types of observational data, which are the administrative claims and electronic (provider-based) health records
- Assessing the performance of the analytical methods for two analysis problems: the monitoring of a defined set of risks and benefits that are relevant to the field of drug outcome research (Health Outcomes of Interest) and identification of non-specified associations
- Studying how the information from these analyses can be successfully used as part of an ongoing program of active drug surveillance, based upon the results of these analytical methods

In the two years since OMOP was launched, it has assembled a central research core of scientists, a program management team responsible for oversight and operations of the research program, technical staff responsible for developing program code for the methods and implementing the research protocols, a Research Lab with access to five central databases, a funded core of five research partners who represent a distributed network of diverse data sources, types, and populations, a dynamic worldwide community of active surveillance methods collaborators, and a public-private partnership with a governance structure for project oversight.

The research team has developed a series of enabling technologies that facilitated the Partnership's methodological research and has made these technologies publicly available for use by the broader research community. This includes the definition and implementation of a Common Data Model (CDM) and a standardized vocabulary to facilitate research across a diverse network of data sources; an infrastructure for a central coordinating center and distributed network of healthcare data partners; standardized procedures for analysis methods developed by a cross-disciplinary community of methods collaborators; systematic tools to characterize and assess data resources for observational analysis; a systematic approach to defining and implementing health outcomes of interest; and tools to generate simulated healthcare data sets to support development of analytical methods. Table 1 provides an inventory of OMOP's key goals and accomplishments that formed the foundation for OMOP's methods program.

Highlighted Accomplishments

Research Laboratory

The OMOP Research Lab accommodates the five central research databases and methods development and testing activities. It consists of two high-end multiprocessor servers, one Oracle server, and thirty-seven terabytes of data. The Research Lab supports the development, testing, and execution of OMOP's

Table 1. Inventory of Goals and Accomplishments

OMOP Key Goals	Accomplishments
Implement a public-private partnership governance model	• 12 Executive Board members, chaired by the FDA and managed by the Foundation for NIH • 21 Advisory Board members • 6 research investigators and Program Management Office
Establish OMOP Research Community	• Built the OMOP Research Lab to accommodate the Common Data Model (CDM) and serve as a central coordinating center • Established a distributed network of 6 data partners • Launched Extended Consortium • Engaged 17 OMOP Methods Collaborators • Hosted OMOP Cup with 60+ participants • Created OMOP website with 1000+ registered users • Held 2009 Symposium with 300+ attendees
Establish a consistent framework to use across disparate observational data sources	• Common Data Model (CDM) • Standardized terminology specifications • CDM reference tables that contain the standardized terminologies and mappings from source vocabularies • ETL specifications for all data partners • GE Centricity and Thomson ETL source code • Generalized ERA logic developer
Develop and test analysis methods within the OMOP Research Lab and other data environments	• Overview of methods (methods “points-to-consider” and inventory matrix) • 12 methods specifications and source code • Observational Medical Data Set Simulator (OSIM) - specification, source code, and data sets
Establish standard data characterization and facilitate comparisons across databases	• Data screen and assessment questionnaires Observational Source Characteristics Analysis Report (OSCAR) Specification and Source Code • Natural History Analysis (NATHAN) Specification and Source Code • Generalized Review of OSCAR Unified Checking for data quality and validation analysis
Implement Health Outcomes of Interest definitions	• Health Outcomes of Interest (HOI) definition process (literature review strategy, evidence table) • HOI process outputs for 10 HOIs • 35 definitions for 10 HOIs • Regularized Identification of Cohorts (RICO) program to implement HOI definitions within CDM

computationally intensive methods across five central databases. Secure communications capabilities provide controlled information exchange with distributed partners. The OMOP team has also established a foundation for a secure cloud-based Research Lab. The OMOP cloud environment provides increased elasticity to meet peak workloads by tapping the computational capacity of up to 250 processing units. OMOP's cloud software is also being made available as open source software.

Framework for Disparate Data

OMOP has developed and made publicly available a CDM, which is a single data schema that can be applied to different databases to establish a consistent structure and format. The CDM is designed to accommodate disparate data types (both administrative claims and electronic health records) and enable the consistent and systematic application of analysis methods to produce comparable results across sources. The CDM provides a framework to assess data quality and creates implementation efficiencies that facilitate rapid analysis. The model has been successfully applied to eleven data resources, and has provided a foundation on which all subsequent OMOP research is based. The work to date has demonstrated that a common model can support active surveillance both in a centralized environment and across a distributed network, and that consistent application of assumptions and rules for data preparation can minimize the variability across sources.

OMOP has delivered technical specifications and programming code to instantiate the CDM, white papers that describe the transformation process followed by each of the data partners, source code for implementing the model for Thomson Reuters and GE Centricity databases, and other standardized procedures for data management that can be applied to any data resources. Together, these deliverables can enable organizations with access to observational data to create their own network of data sources and take advantage of the analysis tools and processes that are being made publicly available throughout the OMOP effort.

A key component of the OMOP CDM is the application of a standardized vocabulary. The standardized vocabulary ensures that methods can be systematically applied to produce meaningful, comparable results across sources, even when databases use different coding schemes to represent the same products and outcomes. The standardized vocabulary provides a mechanism for transforming raw data into standardized data. It also plays a role in searching and querying the transformed data in a CDM database, browsing and navigating the hierarchies of drug classes and conditions, and interpreting the results returned by the analyses. During the development of the specification for the standardized vocabulary, OMOP conducted an evaluation of alternative standards that could be used for active surveillance, such as the potential use of ICD9CM, MedDRA, and SNOMED for classifying adverse events. OMOP developed a database that contains the standards and the mappings necessary to translate codes into these standards. The vocabulary specification and source code, database, and licenses have been made publicly available to the research community.

OMOP Toolkit

- Data Assessments
- Common Data Model (CDM)
- Extract Transform Load (ETL) Process
- Implementation of Standardized Vocabularies (Version 3.0)
- Era Records Development and Integration Programs
- Observational Source Characteristics Analysis Report (OSCAR)
- Descriptive Natural History Analysis (NATHAN)
- Library of Health Outcomes of Interest Definitions
- Methods Library
- Observational Medical Dataset Simulator (OSIM 1) - OSIM 2 in progress

Data Characteristic Tools

Using the OMOP CDM as a foundation, the OMOP team developed several tools and processes to help assess the characteristics of a given data source. The primary tools are the Observational Source Characteristics Analysis Report (OSCAR) and Natural History Analysis (NATHAN). Both OSCAR and NATHAN are available on the OMOP website in the form of a specification document and source code.

OSCAR provides a comprehensive summary of all data available in a CDM. The analysis of the OSCAR results has provided insights into the diversity of data capture across the OMOP community, and the expectations for which medical products and health outcomes may be feasibly explored within an active surveillance system. NATHAN provides a systematic process for exploring the characteristics of subpopulations of interest (such as people exposed to a particular drug or people with a specific condition), to summarize the co-morbidities, concomitant medications, and other clinical observations associated with the subpopulation. Analysis of NATHAN results has helped identify the impact of alternative definitions of health outcomes of interest, and how clinical practice changes over time may influence observed drug effects. The team is actively exploring how these data characteristic tools can contribute to a systematic approach to data quality assessments across an active surveillance network.

Simulated Data

Restricted availability and access to healthcare data has been a barrier to attracting the interest of the methods community in drug safety research. One approach OMOP has taken to meeting this need is the development of simulated data. OMOP (assisted by ProSanos) developed the Observational Medical Data Set Simulator (OSIM) in order to establish a data set with the known "truth" that would facilitate the comparison of performance characteristics within and across methods. The resulting synthetic data sets have been widely used within the OMOP methods development community. The specifications for OSIM 1 as well as the source code and resulting data sets are publicly available to researchers.

With experience gained using the simulated data to support methods development and the OMOP Cup competition, opportunities were identified to improve OSIM 1 and to better reflect complex real-world clinical relationships between drugs and conditions. This led to the development of OSIM 2 that further creates simulated data for hypothetical persons with exposure to real drugs and occurrence of real conditions, but injects hypothetical signals between drugs and conditions to enable method evaluation. The release of the second-generation simulator is scheduled for December 2010.

Analysis Methods

The heart of the OMOP research agenda is the development and assessment of a wide variety of observational analysis methods. To achieve this objective, OMOP released a series of white papers and a call for participation inviting the methods community to collaborate with OMOP to contribute methods to the OMOP Methods Library. In collaboration with methods partners, OMOP has developed fourteen analysis methods for identifying potential drug safety issues. The source code and documentation for twelve of these

Methods Portfolio

- Disproportionality Analysis
- Univariate Self-Controlled Case Series
- Observational Screening
- Multi-Set Case Control Estimation
- Bayesian Logistic Regression
- Case Control Surveillance
- IC Temporal Pattern Discovery
- Case-Crossover
- HSIU Population-Based Method
- Maximized Sequential Probability Ratio Test
- High-Dimensional Propensity Score
- Conditional Sequential Sampling Procedure

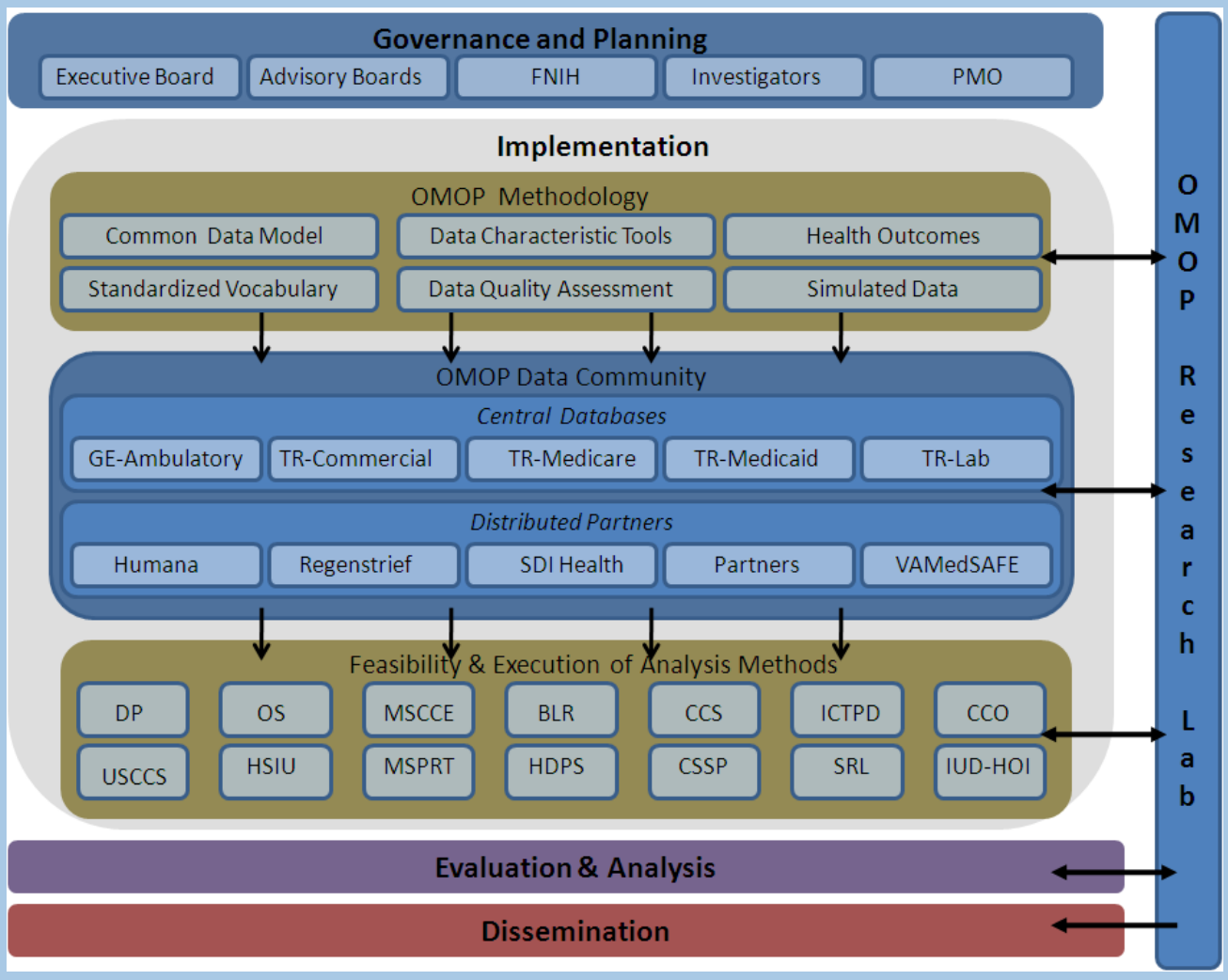
methods have been made publicly available to promote transparency and consistency in research. The remaining two are still undergoing refinement.

Today, the performance of these methods is being evaluated through a series of experiments across the OMOP data community. These analyses should inform the appropriate use of the methods for monitoring Health Outcomes of Interest and identification of non-specified associations. The methods have been developed as standardized procedures that can provide estimates of effects for any drug- and any outcome-allowing organizations to leverage OMOP's tools for their own drug safety analyses against their own observational data sources.

Summary

With this foundation in place (Figure 1), OMOP proceeded into a series of experiments to generate data on the performance of twelve analysis methods. Each of these methods can be configured through parameter settings to determine how well the methods discriminate between the true association and the false associations on the different databases. Today, results from the execution of these methods are being produced from the five data sets in the OMOP Research Lab (centrally) and OMOP's five active Distributed Partners. These experiments have yielded a results data set with 1.2 million observations thus far. The team has begun the process of analyzing results and is on track to complete the original OMOP research agenda during the first quarter of 2011.

Figure 1. The OMOP Operating Model



The Future of OMOP

Building a Sustainable Model

As the Observational Medical Outcomes Partnership (OMOP) approaches its second anniversary as a public-private partnership, the Foundation for the National Institutes of Health (FNIH) is making plans to continue the Partnership's research through 2011 while developing a sustainable and diversified funding base for operations in future years. OMOP was originally launched as a two-year research program that called for the effort to wind down in early 2011. While the research team is on track to complete its work during the first quarter of 2011, the OMOP Executive Board recommended that FNIH continue the partnership. This recommendation reflects the importance of OMOP's research as well as a strong desire to continue the operations of the Research Lab to support methodological research.

With an extension, OMOP will build upon its existing research foundation, continuing its mission to improve drug safety (and benefit) monitoring by:

- Advancing methodological research to explore the performance of methods over time, within populations of interest, and across a broader array of medical products and health outcomes
- Refining and enhancing OMOP's tools and capabilities to translate research into practice
- Sustaining the shared resource (Research Lab) so the research community continues to have access to an open forum for collaborative research

OMOP also intends to diversify its funding sources to include a broader array of stakeholder communities, including federal, healthcare, insurance, biopharmaceutical, information technology, and nonprofit partners.

In 2011, OMOP will continue to advance the field of methodologically research by providing grants for promising ideas, providing data and technical resources to enable research, and by maintaining and applying its talented network of epidemiologists, clinicians, computational scientists, methods developers, and statisticians to methods research. OMOP will continue to pursue research activities that enhance the understanding of observational methods and healthcare data.

The goals for 2011 include (a) additional test cases, (b) periodic data updates to support time-to-detection analytical methods studies, (c) analysis strategies for patient subgroups of interest, (d) design and implement data summaries, visualizations, and analyses to detect database differences, and (e) develop and deliver an improved simulation capability. In order to accomplish these objectives, there is a need to sustain the existing OMOP research community, including the Research Lab, the data and methods collaborators, the central research group, and other community-building activities.

The OMOP Research Laboratory is an important resource and its continuation will serve as a focus for ongoing research across multiple disciplines and stakeholder communities. This unique laboratory, with access to real-world and simulated data, enables a broad research community to develop and test new approaches for improving drug safety analysis capabilities and supports the necessary research to guide development of consistent processes and applications.

The research plan for 2011 builds upon and leverages the current plan. The addition of new test cases, data characterization, and improved data simulation capabilities will further the understanding of the contribution of and synergies among data, technology, methods, and governance required to establish an active surveillance system for pharmaceutical safety using existing observational databases. To gain further value from this information, OMOP is also exploring the potential of using the Research Lab to expand the methods research program into other arenas, such as medical device safety, comparative effectiveness, and healthcare quality. Any activities in these new areas are contingent upon attracting additional stakeholders, funding, and qualified researchers into the partnership.

Follow OMOP

Upcoming Events

ISPE with the Japanese Society for Pharmacoepidemiology
Fifth Asian Conference on Pharmacoepidemiology
October 29-31, 2010
Tokyo, Japan

OMOP 2011 Symposium
January 11, 2011
Washington, DC

OMOP Newsletter

Observational Medical Outcomes Partnership
Foundation for the National Institutes of Health
9650 Rockville Pike
Bethseda, MD 20814-3999
Phone: 301-402-5311
Website: http://omop.fnih.org

Who We Are

OMOP Personnel

Program Management Office:

Thomas Scarnecchia, MS, Executive Director
tscarnecchia@fnih.org

Emily Welebob, RN, MS, Program Manager

Christian Reich, MD, PhD, Program Manager

Research Investigators:

Abraham G. Hartzema, PharmD, MSPH, PhD, FISPE
Professor and Eminent Scholar, Perry A. Foote Chair in Health Outcomes and Pharmacoeconomics; Professor, Department of Epidemiology and Biostatistics, College of Public Health and Health Professions
University of Florida

David Madigan
Professor, Columbia University

Marc Overhage, MD, PhD
Director, Medical Informatics and Research Scientist, Regenstrief Institute, Inc.; Regenstrief Professor of Medical Informatics, Indiana University School of Medicine; CEO and President, Indiana Health Information Exchange

Judith A. Racoosin, MD, MPH
Sentinel Initiative Scientific Lead, Food and Drug Administration

Patrick Ryan, MEng
Johnson & Johnson

Paul Stang, PhD
Senior Director, Epidemiology, Johnson & Johnson Pharmaceutical Research and Development