

# OBSERVATIONAL MEDICAL OUTCOMES PARTNERSHIP

## Letter from the Executive Director

The Observational Medical Outcomes Partnership (OMOP) has completed its originally defined set of research experiments to evaluate empirically the performance of alternative methods on their ability to identify true associations between drugs and outcomes. Released in early 2011, the results highlighted areas for further research, including opportunities for methods enhancement and principled customization of analyses for the attributes of drug-outcome pairs under study.

Today, OMOP is working to evaluate method performance against a larger array of drug-outcome test cases and systematic investigation of both real and simulated data sets. OMOP is focusing on the following five objectives based upon the analyses of methods performance in the initial research findings:

- Improve and expand the evaluation of existing analytical methods.
- Enhance analytical methods.
- Address false positive and false negative drug-health outcomes-of-interest pairs.
- Improve health outcome-of-interest definitions.
- Investigate the impact of grouping drugs or outcomes on method performance.

As you read this newsletter, you will understand how OMOP is undertaking the objectives above with various grants and additional OMOP research investigators.

Also in this newsletter, you will read about the next version of the OMOP Common Data Model (CDM), which is being expanded in collaboration with colleagues at the University of Colorado Denver and the University of California San Diego to accommodate additional data elements and facilitate further research in other use cases. Details regarding the planned enhancements were presented via a conference call on December 2, 2011 and will be posted on the OMOP website. We will have an open comment period addressing questions and will release the CDM update in the beginning of 2012.

You may have already heard from us about EDDIE, the OMOP Epidemiology Design Decision Inventory and Evaluation. Read further to find out how to participate in this unique web-based research study to capture and analyze the potential decisions that epidemiologists, statisticians, clinicians, and other researchers would make when designing a study to determine the association between drugs and potential outcomes of interest. We would like to understand what choices researchers make, what variability in those choices may exist among researchers, what patterns emerge within design choices across drug-outcome pairs, and ultimately how those choices impact the ability to identify true effects, and not identify false findings.

Our research has shown that our work is not done yet, and by looking deeper into our results and expanding our collaborations with other organizations, we can continue to advance the science of observational data analysis.

### In This Issue

**2011 Research Agenda**

**OMOP 2011 Research Projects**

**Extension of the OMOP CDM**

**OMOP Welcomes New  
Research Investigators**

**EDDIE**

**OMOP Publications**

# 2011 Research Agenda

In 2011, the Observational Medical Outcomes Partnership (OMOP) completed its originally defined set of research experiments to evaluate empirically the performance of alternative methods on their ability to identify true associations between drugs and outcomes. To conduct these experiments, OMOP established a data community of ten data partners, containing both administrative claims and electronic health records, and covering over 200 million lives of patient-level data. OMOP also engaged the community of methods developers to implement fifteen alternative analytic approaches for estimating the strength of association between medical product exposure and outcome occurrence. These fifteen methods were executed across the ten databases and applied to fifty-three drug-outcome test cases: nine positive controls and forty-four negative controls. Predictive performance was measured using multiple operating characteristics, including sensitivity, specificity, positive predictive value, and area under ROC curve, by comparing the effect estimates and standard error with the binary classification of the true causal effect.

The empirical findings of the OMOP study suggest that no specific method had superior performance, as each represented different trade-offs between sensitivity and specificity. The initial studies highlighted areas for further research, including opportunities for methods enhancement and principled customization of analyses for the attributes of drug-outcome pairs under study. To achieve these aims, a broader experimental framework was needed to evaluate method performance against a larger array of drug-outcome test cases and systematic investigation of both real and simulated data sets.

The OMOP 2011 research agenda is exploring the interaction among observational data, analytical methods, and drug-outcome relationships. Observational data is being studied through real-world performance of multiple disparate data sets, as well as systematic exploration of simulated data.

Methods development continues to be a focus area, both through new enhancements and innovations, increased parameterization of existing methods, and application of existing tools. Appropriate application of these methods to data for specific drug-outcome associations is being studied empirically through execution of a large array of experiments.

The breadth of test cases has been expanded to allow deeper focus on specific health outcomes-of-interest (acute myocardial infarction, acute liver injury, acute renal failure, and gastrointestinal hemorrhage) with an increased and more robust a priori evaluation of positive and negative controls. The following methods are being applied within the OMOP experiments: cohort, case-control, self-controlled case series, temporal pattern discovery, observational screening, disproportionality analysis, and LGPS/LEOPARD.

Observational screening, temporal pattern discovery, and disproportionality analysis are being executed (in both simulated data and real data) using the implementation developed during the original OMOP experiments. New implementations of the cohort, case-control, and self-controlled case series have been developed by the OMOP research team to improve computational performance, coordinate implementations across disparate methods, and enhance analytic features. LGPS/LEOPARD was modified by Martijn Schuemie of Erasmus University.

Currently, methods are being executed against the five licensed data sets: Thomson MarketScan Commercial Claims and Encounters (CCA), Thomson MarketScan Medicare Supplemental (MDCR), Thomson MarketScan Multi-state Medicaid (MDCD), Thomson MarketScan Lab Supplemental (MSLR), GE Centricity, as well as a series of simulated data sets constructed using OSIM2. We anticipate concluding the execution of the methods experiments during Q1 2012.

# OMOP 2011 Research Projects

The first two years of the Observational Medical Outcomes Partnership's (OMOP) research yielded significant progress towards structured use of existing observational data sources for active surveillance. OMOP established distributed and centralized data access mechanisms with claims and electronic health records data transformed to a common data model. In 2011, one of OMOP's priorities included the refinement of a number of the strategies previously developed. This enrichment to the research began by initiating five grant collaborations in three different categories: enhancing methods, exploring false positives, and exploring existing health outcomes of interest (HOI).

## **Enhancing Methods**

During previously conducted analyses of methods performance, OMOP identified further opportunities to refine methods. These opportunities included improvement of previously developed methods stored in the OMOP methods library and the development of new methods. OMOP is collaborating with Erasmus University Medical Center, Indiana University, and Uppsala Monitoring Centre to explore these opportunities.

### ***Erasmus University Medical Center: Detection of Long-Term Adverse Drug Reactions in Electronic Healthcare Data***

A substantial group of adverse drug reactions (ADRs) is likely to have a much longer time between first exposure and adverse event occurrence, and the risk could actually be increasing with cumulative exposure. Martijn Schuemie, PhD, Assistant Professor, Medical Informatics Department, Erasmus University Medical Center in Rotterdam, Netherlands, leads this work, stating that "...the availability of large amounts of longitudinal observational healthcare data with long follow-up now puts us in a unique position where we would be able to pick up these types of ADRs, but appropriate signal detection screening methods are lacking."

The primary objective of this research is to develop methods that can detect long-term ADRs, using simulated data and real data. Methods will be

developed to detect ADRs that are correlated with cumulative exposure, and to identify automatically groups of patients that can be used as comparison to eliminate selection bias as much as possible.

### ***Indiana University: Cohort Design Enhancements***

In the initial method collaboration with OMOP, Xiaochun Li, PhD, Department of Biostatistics, Indiana University School of Medicine, developed a simple cohort approach called HSIU to screen associations between medications and non-specified conditions in very large clinical and administrative databases. "We learned through the OMOP experiments that analytic results are extremely sensitive to design features of methods," states Dr. Li.

With this new research, the Indiana University team is exploring cohort-design methods, and will develop a program that will allow the study of outcomes with flexible design features such as baseline duration, options for follow up, and options for user-specified covariates, eligibility, and outcomes. This program will also allow investigators to choose between new user or prevalent user designs for a specific investigation of a drug and an outcome. Dr. Li indicates, "To better control the balance between sensitivity and specificity and to improve the overall accuracy through bias reduction, we anticipate that the implementation of a series of options allowing for flexibility in the cohort study design and inclusion of user-defined covariates will help to achieve better classification of drug-outcome associations."

The updated cohort program will be executed against the Regenstrief Institute implementation of the OMOP Common Data Model (CDM) and will be made available to run on databases with the OMOP CDM format.

### ***Uppsala Monitoring Centre: IC Temporal Pattern Discovery Improvements and In-Depth Analysis of Clinical Outcomes of Interest***

The IC Temporal Pattern Discovery method (IC-TPD) was implemented during the initial OMOP initiative. This current research, led by Niklas Norén, PhD, Uppsala Monitoring Centre, Manager, Research Department, will improve upon the algorithm based

upon the previous performance of IC-TPD. Dr. Norén and his team are conducting an in-depth investigation of the temporal patterns for false positive and negative findings, with the aim of understanding potential limitations of the current methodology and proposing further improvements.

### **Exploring False Positives**

Within the OMOP experiment, certain drug-HOI pairs that have a known association were consistently missed by a wide range of methods throughout multiple databases (false negatives), whereas some negative control drug-HOI pairs were consistently identified as showing an association (false positives). Initial investigation of these unexpected findings did not reveal adequate explanations. A number of hypotheses were generated but were not systematically explored. To address this problem, the OMOP research team and collaborators have been examining the false negatives and false positives in detail.

### **Massachusetts General Hospital: Validating OMOP Results with Reproducible Detailed Data Investigation**

The primary research objective of this work is to investigate three of the false positive drug-HOI pairs that arose during the initial OMOP research. The three false positives being studied by Shawn Murphy, MD, PhD, Medical Director, Research Computing for Partners HealthCare System, include: antibiotics and acute renal failure, typical antipsychotics and GI ulcer hospitalization, and warfarin and hip fracture. The Informatics for Integrating Biology and the Bedside (i2b2) workbench will be used to explore these false positive results. Dr. Murphy and his team are developing a standardized approach for looking deeply into the data in order to understand better how to correctly identify true associations in the coded data, which, importantly, will be enriched by the addition of text notes from the patient record. They have also built a process that transforms the OMOP CDM and Ontology to the i2b2 platform, allowing for the investigation of the three drug-HOI false positive pairs.

### **Improving Existing HOIs**

Definitions for measurement of HOIs in observational data are a critical element to bolster confidence in output of the active surveillance process. During OMOP research, HOIs were defined

using an extensive systematic literature review followed by an expert panel review to create multiple HOI definitions. It is difficult to imagine investing this level of effort for every HOI needed, and confidence in the HOI definitions is still limited with this approach. Instead, OMOP will collaborate with Auburn University to develop a structured process for measuring and interpreting HOIs.

### **Auburn University: Developing a Structured Process for Measuring and Interpreting HOIs in the OMOP CDM**

A public-private partnership between Auburn University and Hewlett Packard (HP) Labs is furthering the OMOP HOI research via their study. The Auburn team includes Richard Hansen, PhD (PI), Brent Fox, PharmD, PhD, and Joshua Hollingsworth, PharmD, from the Department of Pharmacy Care Systems, and David Mark Carpenter, PhD, from the Department of Mathematics and Statistics. The HP Labs team is led by Michael Gray, PhD.

Their research will investigate a subset of the OMOP HOIs, starting with acute liver injury. The overall goal of this work is to minimize false positive cases through better measurement of the HOI. “We are conducting refinement of HOI definitions through clinical review of cases generated from de-identified patient-level data sources, creating training data sets through expert classification of true cases, using predictive modeling to further refine HOI definitions, and comparing methodological performance based on probability thresholds,” said Dr. Hansen. “We are using an expert panel to reach consensus on whether identified cases are likely true cases, and then using this case review process to improve HOI definitions.”

In addition to these five grants, OMOP has extended work with the Department of Veterans Affairs Center for Medication Safety to conduct additional analyses on prior results for a select number of the DOI-HOI pairs. “We are very pleased to have the opportunity to work with such experienced investigators and know that the findings from their work will contribute greatly to OMOP research goals and the broader research community to improve drug safety”, said Thomas Scarnecchia, OMOP Executive Director.

## Extension of the OMOP CDM

The Observational Medical Outcomes Partnership's (OMOP) Common Data Model (CDM) was designed to accommodate data from a variety of medical databases, such as electronic health records and claims databases to rapidly assess active drug safety surveillance using data already collected for administrative or clinical care purposes. In the original OMOP project, ten different observational data sets were translated and stored in OMOP CDM databases, which allowed for a number of analytic methods to be executed across all of the databases. Converting multiple disparate databases to a CDM allowed investigators to write and test the analytic scripts once and then run them on all of the databases with minimal modification. The OMOP CDM showed to be a promising collaborative infrastructure for active drug safety surveillance and, in 2011, it was decided that this same infrastructure might support broader applications of comparative effectiveness research, beyond drug safety studies.

One of OMOP's tenets is to establish a shared resource (tools and capabilities) so that the broader research community can collaboratively advance the science of comparative and surveillance studies. Lisa Schilling, MD, MSPH, at the University of Colorado School of Medicine, Colorado Health Outcomes Program, is the Principal Investigator on the AHRQ-funded project, entitled "The Scalable Architecture for Federated Therapeutics Inquiries Network (SAFTINet)." SAFTINet is a project to build a distributed research network, requiring a CDM. Dr. Schilling and Michael Kahn, MD, PhD, SAFTINet Co-Investigator and Informatics Lead, approached OMOP to extend collaboratively the OMOP CDM to better support SAFTINet's goal to create a CDM to support comparative effectiveness research.

SAFTINet's multi-state project will allow researchers, health policy experts, payers, and clinicians to better understand the impact of a wide variety of healthcare interventions on health

outcomes for minority, underserved, and socioeconomically disadvantaged populations by supporting observational comparative effectiveness research. SAFTINet will both leverage and extend the established governance and technologic capabilities of the Distributed Ambulatory Research in Therapeutics Network (DARTNet). The University of Colorado team also engaged collaborators from the University of San Diego's SCALable National Network for Effectiveness Research (SCANNER) team including Co-PIs: Drs. Lucila Ohno-Machado and Aziz Boxwala, Daniella Meeker from Rand Corporation, and Omolola Ogunyemi from Charles Drew University of Medicine and Science, to ensure the planned OMOP extensions meet the needs of a broad comparative effectiveness research (CER) community.

Dr. Schilling relates, "We are building a distributed research network combining ambulatory and inpatient clinical data and Medicaid claims and eligibility data for clinical and research purposes—we did not want to recreate a common data model, so our research team choose to use OMOP's CDM and add tables for additional data types to accommodate our research."

Dr. Kahn reinforced Dr. Schilling's thoughts, "We are excited to be part of the OMOP community and contribute to the next version of the OMOP CDM by adding tables to capture cost information, as well as information about healthcare providers within the healthcare systems in the network. With our focus on data from electronic medical records, we will contribute back to the OMOP community new data transformation documentation that highlights issues with protected health information that were not relevant to the initial data sources, thus extending the reach of OMOP investigations to include these new rich clinical data sources."

Watch the OMOP website <http://omop.fnih.org/> for the comment period of the next CDM version.

## OMOP Launches a Unique Study

The Observational Medical Outcomes Partnership (OMOP) has launched a unique study entitled, “Epidemiology Design Decision Inventory and Evaluation” (EDDIE). The purpose of the study is to capture and analyze the potential decisions that epidemiologists, statisticians, clinicians, and other researchers would make when designing a study to determine the association between drugs and potential outcomes of interest (for example, ACE inhibitors and angioedema). The objective is to better understand what choices researchers would make, what variability in those choices may exist among researchers, what patterns emerge within design choices across drug-outcome pairs, and, ultimately, how those choices impact the ability to identify true effects and not identify false findings.

OMOP seeks volunteers from epidemiological, statistical, and clinical research backgrounds to participate in this deep dive into epidemiology study design. The survey contains a series of questions about how you would design an epidemiology evaluation study for specific drug-outcome pairs and specific decisions about how that design would be implemented. A web-based survey, this study can be completed within your own setting and on your own timetable. Detailed information on how to participate is located on the OMOP website at <http://omop.fnih.org>. Questions can be addressed to Dr. Patrick Ryan, OMOP Research Investigator, at [ryan@omop.org](mailto:ryan@omop.org), or Emily Welebob, OMOP Program Manager, at [welebob@omop.org](mailto:welebob@omop.org). The study will end on December 31, 2011.

## OMOP Publications

(Ahead of print article) [Available online through the journal website](#)

Overhage JM, Ryan PB, Reich CG, et al. Validation of a common data model for active safety surveillance research. *J Am Med Inform Assoc* (2011). doi:10.1136/amiajnl-2011-000376.

(Ahead of print article) [Available online through the journal website](#)

Zorych I, Madigan D, Ryan P, and Bate A. Disproportionality methods for pharmacovigilance in longitudinal observational databases. *Stat Methods Med Res* 0962280211403602, first published on August 30, 2011 as doi:10.1177/0962280211403602.

Madigan D, Ryan P. What can we really learn from observational studies? The need for empirical assessment of methodology for active drug safety surveillance and comparative effectiveness research. *Epidemiology*. 2011;22:629–631.

Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, et al. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. *Ann Intern Med*. 2010 Nov 2;153(9):600-6.

Ryan PB, Welebob E, Hartzema AG, Stang PE, Overhage JM. Surveying US observational data sources and characteristics for drug safety needs. *Pharm Med*. 2010; 24 (4): 231-238.

# OMOP Welcomes New Research Investigators

Since the initiation of the Observational Medical Outcomes Partnership (OMOP), principal investigators (PIs) have played a critical role in the guidance and completion of OMOP research. Investigators have been from the government, pharmaceutical industry, and academia. OMOP's research needs have further developed leading to the opportunity to expand the PI group. In addition to the current OMOP PIs, OMOP is pleased to welcome three new PIs to lead and guide the OMOP 2011 research experiments.

**Dr. William DuMouchel, PhD**, is a chief statistical scientist at Oracle Health Sciences, Tuscon, Arizona. He previously was the Chief Statistical Scientist of Lincoln Technologies, Inc., and a senior statistician at AT&T Labs--Research. From 1987 to 1992, he served as the Chief Statistical Scientist at BBN, where he was responsible for the statistical aspects of their computer advisory systems for design of experiments and data analysis called RS/Discover and RS/Explore. Dr. DuMouchel has been on the faculties of the University of California at Berkeley, the University of Michigan, MIT, and at Columbia University. He is the inventor of the empirical Bayesian data-mining algorithm known as GPS and its successor MGPS, which have been applied to the detection of safety signals in databases of spontaneous adverse event reports. He is a frequent speaker at industry conferences on data mining and statistical methods. He received a PhD in Statistics from Yale University in 1971.

**Dr. Marc Suchard, MD, PhD**, is a professor in the Departments of Biomathematics, Biostatistics,

and Human Genetics at UCLA's David Geffen School of Medicine and School of Public Health. After earning his B.S. in biophysics (with distinction) from the University of California, Berkeley, in 1995, Dr. Suchard spent two years at Oxford University as a British Marshall Scholar. He then continued his studies at the University of California, Los Angeles, earning a PhD in biomathematics in 2002. He earned his MD degree also from UCLA. Dr. Suchard has authored or coauthored many papers. With OMOP, Dr. Suchard is using his expertise in massive parallelization in computational statistics through graphics processing units (GPUs) to improve OMOP's methodology research and he is also contributing general methodological expertise.

**Dr. Martijn Schuemie, PhD**, is an assistant professor at the Medical Informatics Department of the Erasmus University Medical Center in Rotterdam, the Netherlands. He has a master's degree in business economics (Information Management), and a PhD in informatics. In the past, he has worked extensively on text mining, including the clustering of large sets of documents. He is currently working on pharmacovigilance methods, and is heading the comparison of signal detection methods in the EU-ADR project. He designed and developed the distributed data analysis software Jerboa®, which is being used in a wide range of European projects, including EU-ADR12, SOS, and VAESCO. In the OMOP Cup, he achieved the highest score using newly developed methods. These methods are currently being evaluated in both EU-ADR and the OMOP experiment.

## OMOP Newsletter

©2011 Foundation for NIH, All Rights Reserved.

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

