Individuals and Organizations Completing Research in the IMEDS Lab
OMOP Investigator Group

Proposed research by OMOP Investigator Group –12/12/13

The OMOP Investigator Group consists of David Madigan PhD, Professor and Chair of the Department of Statistics, Marc Suchard MD, PhD, Professor of Biomathematics at UCLA, Marc Overhage MD, PhD, Chief Medical Informatics Officer at Siemens Health Services, William DuMouchel PhD, Chief Statistical Scientist at Oracle Health Sciences, Abraham Hartzema, PharmD, MSPH, PhD, FISPE, Professor and Eminent Scholar Pharmaceutical Outcomes and Policy, University of FL Paul Stang PhD, Vice President of Epidemiology at Janssen Research and Development, Patrick Ryan PhD, Head of Epidemiology Analytics at Janssen Research and Development, Martijn Schuemie PhD, Associate Director, Epidemiology Analytics at Janssen Research and Development, and Christian Reich MD, PhD, Global Head of Discovery Informatics at AstraZeneca. Access is requested for Ryan, Schuemie, Reich, Madigan, Suchard, Hartzema and Overhage. All except Hartzema and Overhage already have accounts.

Research Objectives and Proposed Approaches
This project will focus on software and infrastructure to support post-marketing evidence generation by creating summary statistics for associations between all combinations of medical products and outcomes, to be examined in the context of relevant evidence such as the literature and spontaneous reporting data to improve our understanding of how to interpret the safety and comparative effectiveness data derived from observational (secondary) data. We do not differentiate safety surveillance from comparative effectiveness, and do not intend to apply different analytical methods beyond what has already been explored within OMOP's past research. Our goal is the enable the systematic exploration of the effects of medical products through a comprehensive causal assessment framework, in part by providing standardized effect estimates using traditional study designs, including case-control, cohort, and self-controlled methods. The cohort method, when applied using an active comparator as a reference group, can be interpreted as either a safety study or comparative effectiveness study, and we intend to use the results of our analysis to allow further exploration of the impact of choice of comparator on effect estimation and method predictive accuracy.

Aim 1: Create test cases for a comprehensive number of outcomes. We will use the observational data to identify the drugs and outcomes are observable and have sufficient power to be used for analysis. Where we can, we will use information from product labels, literature, spontaneous adverse event reporting and existing biomedical ontologies to identify drugs that can serve as controls for outcomes to be used to calibrate *p*-values. Negative controls are not specific to safety or effectiveness because we are simply looking for drug-outcome pairs with 'no effect'.

Method	Abbreviation	Total parameter combinations
Cohort	CM	126
Case-control	CC	384
Self-controlled case series	SCCS	560
Observational screening	OS	54
Temporal pattern discovery	ICTPD	42
Disproportionality analysis	DP	48
Longitudinal Gamma Poisson Shrinker	LGPS	32

Aim 2: Characterize the association between drug-outcome combinations. We measure the association between all drugs and all outcomes using methods developed for the 2011-13 OMOP experiments (see table) and selected analysis parameter choices (analyses) to minimize the number of analyses required based on critical review and analysis of the OMOP experiment results to date. We intend to estimate the effects of medical products by exploring the temporal association between medical product exposure and outcome incidence. Those outcomes include the onset of conditions and utilization of health services for the

diagnosis and treatment of those conditions. We will restrict the analysis to those medical products and outcomes that are observable in the databases, and to outcomes for which we can establish a sample of negative controls (drugs for which we do not believe there is an association with the outcome) which can be used to calibrate p-values from estimates across all drugs which are not negative controls (Aim 1).

Aim 3: Create summary statistics derived from the patient-level observational data that can be combined with other information sources in a publicly available risk identification and comparative effectiveness system. We will utilize the results of this project to provide association data and specifically variation in the association estimates across databases, methods and analysis parameters that provide insight into what stakeholder can infer from observational data.

Impact
Associations derived from observational data may provide important information to support decision making by a variety of stakeholders. Creating part of the infrastructure necessary to support an active medical product safety surveillance system will provide both practical experience and insights from usage of these data.

Experience
The researchers for whom access is requested are all well qualified for working with this type of data having established the lab as part of the OMOP project and used it extensively to study methods performance.

Data Requirements
We request access to all five of the data sets currently available in the lab and to any that may be added as the work progresses.

Funding Sources
The OMOP team has no external sources of funding to support this effort, and seeks support from IMEDS to cover the IMEDS Lab expense. We expect to run the analysis using a collection of Windows 32 Ora/RedShift images, using High-CPU Medium type machines with 1TB of temporary storage. These instances are estimated to cost $0.473/hr. Assuming we use 10 instances for the full 2-year period, the total computing cost would be expected to be $82,869 over the course of the project, and can be capped to not exceed $100,000. The analysis can be conducted on time and within budget, assuming that all IMEDS Lab databases continue to be centrally maintained in the RedShift 8-node environment, as was supported in 2013 for analyses demonstrated at the OMOP Symposium. Additional personnel support from IMEDS technical staff is expected to be needed to configure OMOP analysis methods to execute within the RedShift environment, since earlier implementations were developed to run in SAS/Oracle.

Duration
We anticipate completing this work will require 24 months.

References

Stang, Paul E., Patrick B. Ryan, Judith A. Racoosin, J. Marc Overhage, Abraham G. Hartzema, Christian Reich, Emily Welebob, Thomas Scarnecchia, and Janet Woodcock. "Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership." *Annals of internal medicine* 153, no. 9 (2010): 600-606.

Overhage, J. Marc, Patrick B. Ryan, Christian G. Reich, Abraham G. Hartzema, and Paul E. Stang. "Validation of a common data model for active safety surveillance research." *Journal of the American Medical Informatics Association* 19, no. 1 (2012): 54-60.

Stang, Paul E., Patrick B. Ryan, Stacie B. Dusetzina, Abraham G. Hartzema, Christian Reich, J. Marc Overhage, and Judith A. Racoosin. "Health outcomes of interest in observational data: issues in identifying definitions in the literature." *Health Outcomes Research in Medicine* 3, no. 1 (2012): e37-e44.

Hartzema, Abraham G., Christian G. Reich, Patrick B. Ryan, Paul E. Stang, David Madigan, Emily Welebob, and J. Marc Overhage. "Managing data quality for a drug safety surveillance system." *Drug Safety* 36, no. 1 (2013): 49-58.

Reich, Christian, Patrick B. Ryan, Paul E. Stang, and Mitra Rocca. "Evaluation of alternative standardized terminologies for medical conditions within a network of observational healthcare databases." *Journal of biomedical informatics* 45, no. 4 (2012): 689-696.

Murray, Richard E., Patrick B. Ryan, and Stephanie J. Reisinger. "Design and validation of a data simulation model for longitudinal healthcare data." In *AMIA Annual Symposium Proceedings*, vol. 2011, p. 1176. American Medical Informatics Association, 2011.

Hartzema, Abraham G., Christian G. Reich, Patrick B. Ryan, Paul E. Stang, David Madigan, Emily Welebob, and J. Marc Overhage. "Managing data quality for a drug safety surveillance system." *Drug Safety* 36, no. 1 (2013): 49-58.

Ryan, Patrick B., Martijn J. Schuemie, Emily Welebob, Jon Duke, Sarah Valentine, and Abraham G. Hartzema. "Defining a reference set to support methodological research in drug safety." *Drug Safety* 36, no. 1 (2013): 33-47.

Ryan, P. B., D. Madigan, P. E. Stang, M. J. Schuemie, and G. Hripcsak. "Medication-Wide Association Studies." *CPT: Pharmacometrics & Systems Pharmacology* 2, no. 9 (2013): e76.

Ryan, Patrick B., Martijn J. Schuemie, Susan Gruber, Ivan Zorych, and David Madigan. "Empirical performance of a new user cohort method: lessons for developing a risk identification and analysis system." *Drug Safety* 36, no. 1 (2013): 59-72.

Li, Xiaochun, Siu Hui, Patrick Ryan, Marc Rosenman, and Marc Overhage. "Statistical visualization for assessing performance of methods for safety surveillance using electronic databases." *Pharmacoepidemiology and drug safety* (2013).

Ryan, Patrick. "Using Exploratory Visualization in the Analysis of Medical Product Safety in Observational Healthcare Data." In *A Picture is Worth a Thousand Tables*, pp. 391-413. Springer US, 2012.

Ryan, Patrick B., Martijn J. Schuemie, and David Madigan. "Empirical performance of a self-controlled cohort method: lessons for developing a risk identification and analysis system." *Drug Safety* 36, no. 1 (2013): 95-106.

Ryan, Patrick B., Martijn J. Schuemie, and David Madigan. "Empirical performance of a self-controlled cohort method: lessons for developing a risk identification and analysis system." *Drug Safety* 36, no. 1 (2013): 95-106.

Schuemie, Martijn J., David Madigan, and Patrick B. Ryan. "Empirical performance of LGPS and LEOPARD: lessons for developing a risk identification and analysis system." *Drug Safety* 36, no. 1 (2013): 133-142.

Simpson, Shawn E., David Madigan, Ivan Zorych, Martijn J. Schuemie, Patrick B. Ryan, and Marc A. Suchard. "Multiple self-controlled case series for large-scale longitudinal observational databases." *Biometrics* (2013).

Madigan, David, Patrick B. Ryan, and Martijn Schuemie. "Does design matter? Systematic evaluation of the impact of analytical choices on effect estimates in observational studies." *Therapeutic Advances in Drug Safety* 4, no. 2 (2013): 53-62.

Reich, Christian G., Patrick B. Ryan, and Marc A. Suchard. "The impact of drug and outcome prevalence on the feasibility and performance of analytical methods for a risk identification and analysis system." *Drug Safety* 36, no. 1 (2013): 195-204.

Ryan, Patrick B., Martijn J. Schuemie, and David Madigan. "Empirical performance of a self-controlled cohort method: lessons for developing a risk identification and analysis system." *Drug Safety* 36, no. 1 (2013): 95-106.

Madigan, David, Patrick B. Ryan, Martijn Schuemie, Paul E. Stang, J. Marc Overhage, Abraham G. Hartzema, Marc A. Suchard, William DuMouchel, and Jesse A. Berlin. "Evaluating the Impact of Database Heterogeneity on Observational Study Results." *American journal of epidemiology* (2013).

Ryan, Patrick, Marc A. Suchard, Martijn Schuemie, and David Madigan. "Learning from Epidemiology: Interpreting Observational Database Studies for the Effects of Medical Products." *Statistics in Biopharmaceutical Research*just-accepted (2013).

DuMouchel, William, Patrick B. Ryan, Martijn J. Schuemie, and David Madigan. "Evaluation of disproportionality safety signaling applied to healthcare databases." *Drug Safety* 36, no. 1 (2013): 123-132.

Madigan, David, Martijn J. Schuemie, and Patrick B. Ryan. "Empirical performance of the case–control method: lessons for developing a risk identification and analysis system." *Drug Safety* 36, no. 1 (2013): 73-82.