Getting Start Guide for Exploring the OMOP 2011/2012 Experiment Results

Patrick Ryan

28 September 2012

This document is intended for those interested in working directly with the OMOP Experiment Results, which are all publicly available through the OMOP website.

- Read a brief summary on the OMOP 2011/2012 experiment and findings, read: http://omop.fnih.org/sites/default/files/OMOP%20summary%20July%202012.pdf.
- For further detail, review the presentations from the OMOP 2012 Symposium: http://omop.fnih.org/2012SymposiumPresentations
- Go to: http://omop.fnih.org/Research
    - Download the 'OMOP 2011-2012 Experiment Method Reference'
        - This is an Excel file that contains descriptions of the parameters used within each study design that make up the unique ANALYSIS_ID
        - You will use this file as reference when exploring differences in performance across methods and design parameters
    - Download the 'OMOP 2011-2012 Experiment Method Results'
        - The download is a .zip file containing one .csv file, which holds all results from all analyses conducted across 5 real databases for 399 test cases (positive controls and negative controls) for 4 Health Outcomes of Interest (Acute myocardial infarction, acute renal failure, acute liver injury, upper gastrointestinal bleeding).
        -
- To learn more about the 5 real databases, refer to: http://omop.fnih.org/sites/default/files/Data%20Sources%20for%20OMOP%202011-2012%20Research_FINAL.pdf
- To learn more about the 7 methods applied to the data, refer to: http://omop.fnih.org/sites/default/files/Methods%20in%20OMOP%202011_2012%20Research.pdf
    - For further detail about the specific implementation of each method, refer to: http://omop.fnih.org/MethodsLibrary
- To learn more about the process to develop the reference set of positive controls and negative controls, refer to: http://omop.fnih.org/sites/default/files/OMOP%202011-2012%20Reference%20Set_FINAL.pdf

Data Dictionary for OMOP_2011_METHOD_RESULTS

Variable	Variable description
METHOD_ABBR	Abbreviation for the method used to generate the analysis. Allowable values: CM – Cohort Method CC – Case control DP – Disproportionality Analysis ICTPD – Information Component/Temporal Pattern Discovery LGPS – Longitudinal Gamma Poisson Shrinker OS – Observational screening SCCS – Self-controlled case series
ANALYSIS_ID	Unique identifier for the specific combination of parameters used within the method. Use the Method Reference spreadsheet to look up the specific parameters associated with each analysis_id.
DRUG_CONCEPT_ID	Unique identifier for drug studied. Drug are evaluated at the RxNorm ingredient level. These CONCEPT_IDs are available for lookup in the OMOP Vocabulary v4 (http://omop.fnih.org/CDMvocabV4), though the specific CONCEPT_NAME is made available in this file for your convenience.
DRUG_CONCEPT_NAME	The CONCEPT NAME associated to the DRUG_CONCEPT_ID.
CONDITION_CONCEPT_ID	Unique identifier for outcome studied. These CONCEPT_IDs are available for lookup in the OMOP Vocabulary v4 (http://omop.fnih.org/CDMvocabV4), though the specific CONCEPT_NAME is made available in this file for your convenience. These CONCEPT_IDs reflect the custom Health Outcome of Interest definitions developed by OMOP. The specific coding algorithms used for each HOI definition is available at: http://omop.fnih.org/HOI.
CONDITION_CONCEPT_NAME	The CONCEPT NAME associated to the CONDITION_CONCEPT_ID.
LOGRR	Natural log of relative risk estimate. All methods yield estimates of the strength of association, and where possible have been transformed to a common scale to facilitate comparisons across methods and databases. Note, methods may produce effect estimates on different scales (e.g. odds ratio, incidence rate ratios). Please refer to each method implementation for specific details.
SELOGRR	Standard error of the log relative risk. A measure of variance around the strength of association, as estimated by each method.
RR	Relative risk. The strength of association measure, which is the exponentiation of LOGRR.
LB95RR	Lower bound of 95% confidence interval around the relative risk estimate. This is estimated using the standard normal approximation, using the transformed standard error and relative risk.

Variable	Variable description
UB95RR	Upper bound of 95% confidence interval around the relative risk estimate. This is estimated using the standard normal approximation, using the transformed standard error and relative risk.
GROUND_TRUTH	The classification of the drug-outcome pair. GROUND_TRUTH can be used for measuring predictive accuracy, such as Area under ROC Curve (AUC), or stratifying outcomes (such as measuring mean squared error for negative controls, assuming true RR = 1). Allowable values: 1 = Positive Controls: Drug-outcome pairs that satisfy the following criteria: • Event listed in Boxed Warning or Warnings/Precautions section of active FDA structured product label • Drug listed as 'causative agent' in Tisdale et al, 2010: "Drug-Induced Diseases" • Literature review identified no powered studies with refuting evidence of effect 0 = Negative Controls: Drug-outcome pairs that satisfy the following criteria: • Event not listed anywhere in any section of active FDA structured product label • Drug not listed as 'causative agent' in Tisdale et al, 2010: "Drug-Induced Diseases" • Literature review identified no powered studies with evidence of potential positive association
MIN_DETECTABLE_RR	Minimum detectable relative risk – a measure of data availability analogous to power and sample size considerations, appropriate in the context of retrospective data. We calculate the minimum detectable relative risk by estimating the prevalence of the drug exposure and outcome occurrence in each age decile-by-gender strata, and calculating the number of expected events assuming independence, then aggregating across strata. The composite 'expected events' is then used to estimate the minimum detectable relative risk, assuming a standard cohort design, with Type I error = 0.05 and Type II error = 0.80. The specific approximation applied is discussed in Armstrong AJE 1987. In the OMOP experiment, minimum detectable relative risk was used to stratify test cases on the basis of which seem to have sufficient sample to be reliably studied in a given database.