REal World Assessment and Research of Drugs (REWARD): presenting an open-source package for Population-level effect estimation at the scale of all outcomes by all exposures

James P. Gilbert, Justin Bohn, Chris Knoll, David M. Kern

Background
The application of observational health data in the context of drug safety evaluation has been well studied to have strong methodological grounding (1). In previous work (2–5) we have described the application of the REWARD platform for the discovery of unknown benefits of existing medications. In this work we present the fully open-source application1, built on top of OHDSI HADES software, for use on top of any OMOP Common Data Model (CDM) compliant database.

Methods
The formal requirement for REWARD is an OMOP CDM, and an analysis environment capable of running OHDSI R packages. In addition, REWARD requires a database resource to store aggregated results2. All exposures are based on a new user cohort design, these can be at the RX Norm Ingredient level, ATC classifications or customized phenotype algorithms. Included patients are indexed by exposure date and require 365 days of prior exposure. Automated disease outcomes are based on SNOMED codes, which allows for transportability between different data sources. These automated definitions are useful for empirical calibration of outcomes and come in three varieties: an appearance of at least a single code in a patient’s history, two or more codes in a patient’s history or a code during an inpatient visit. In addition (and for significantly more reliability in patient capture) well defined phenotypes can be included (e.g., the OHDSI phenotype library).

The basis of REWARD is the self-controlled cohort (SCC) study design (6) which overcomes time invariant confounding by allowing patients to act as their own controls. This method is also appealing as it allows satisfactory performance at the scale of all exposures by all outcomes. These, uncalibrated, results are blinded population level estimates between exposure and outcome pairs and are not used until empirically calibrated estimates (7) can be generated when investigating specific exposures or outcomes of interest, negative controls are selected automatically using the common evidence model (CEM) (8) using the CemConnector package3. At this point, R shiny dashboards are generated to allow exploration based on specific research questions including meta-analysis across selected data sets.

Results
REWARD is a fully open-source application the outline of which is shown in Figure 1. One feature of REWARD is that is designed to allow execution in multiple environments satisfying a range of criteria. Execution requires a central database for aggregation of results and storage of cohort references. These

1 Available at https://github.com/OHDSI/Reward
2 Currently, only PostgreSQL is supported as a backend but in principle any database compliant with OHDSI tools could be used.
3 Available at https://github.com/OHDSI/CemConnector
references can be exported, and the execution completed using flexible workflows using the R targets package (9) (see Figure 2).

To date, REWARD has been applied to four US administrative claims databases described fully in (5), in addition to the Japan Medical Data Center (JMDC). 197,032,729 observed effect estimates have been computed across 3,384 exposure and 66,331 outcome definitions.

In Figure 3 we demonstrate the empirical calibration of effect estimates using automatically selected negative controls for Ibuprofen and Warfarin. Here we note that there is a strong left bias apparent in the study design, indicating the need for confirmation from multiple data sources and usage of diagnostic steps in studies to rule out false positives. REWARD provides an R shiny app utility to explore study bias before further evaluating results.

**Conclusion**

We have presented REWARD, a large scale, open-source platform for the evaluation and exploration of population level effect estimates. This platform has been applied to several notable studies, providing valuable evidence for finding the unexpected benefits of existing medications. However, it should be noted that several significant limitations exist within this framework. Most notably, the SCC method shows protective bias. This is corrected with empirical calibration; however, many estimates have elevated levels of systematic error. In the examples explored here, this bias is likely due to the wide range of uses for prescriptions of common drugs. Consequentially, caution is required when evaluating effect estimates generated at this scale.

**Figure 1:** REWARD system outline. The system can be broken into 4 component parts: The management database, which centralizes aggregated results and cohort references, the execution package, which generates raw effect estimates, the data aggregation steps and, the Shiny web app.
Figure 2: REWARD targets (9) execution workflow for generating PLE estimates. Workflow steps can be modified with ease and can be scaled for parallel execution across clusters.

Figure 3: Calibration fan plots of automatically selected negative controls for Ibuprofen (left) and Warfarin (right) on IBM Commercial Claims and Encounters (CCAE) database shows extreme protective bias resulting in an expected absolute systematic error of 0.703 and 0.532. Estimates below the dashed line (gray area) have $p < 0.05$ using traditional $p$-value calculation. Estimates in the orange areas have $p < 0.05$ using the calibrated $p$-value calculation and the red band indicates the 95% credible interval of the boundary of the orange area. Blue dots indicate negative controls.
References


