***[Note: All submissions must be in PDF format. Failure adhere to the format requirements may result in rejection of your submission without review]***

| | |
|---|---|
| Name: | Jenna Marie Reps |
| Affiliation: | Janssen |
| Email: | jreps@its.jnj.com |
| Presentation type (select one): | Poster |

# Utilizing the OHDSI collaborative network for large-scale prognostic model validation

**Jenna M. Reps, PhD[1], Patrick B. Ryan, PhD[1], Peter Rijnbeek, PhD[2], Martijn J. Schuemie, PhD[1]**
**[1]Janssen Research and Development, Raritan, NJ; [2]Erasmus MC, Rotterdam, The Netherlands;**

**Abstract**

*A common criticism of prognostic models is their lack of external validation. This can limit the clinical impact models make because there is greater uncertainty around how well a model will generally perform. The OHDSI network presents the perfect opportunity to implement large-scale, insightful prognostic model validation by enabling researchers to easily share their models with other collaborators in order to evaluate the model across a number of datasets available to the community. In this paper, a framework for implementing the large-scale model validation is proposed and tested. The results show that applying a prognostic model across a large number of datasets can help ensure the model is capable of generalizing to new data or highlight the model's limitations. This is an important step towards developing models that are most likely to have a clinical impact.*

**Introduction**

Prognostic models aim to predict the risk of an outcome occurring during some specified future time period based on a patient's current medical state. These models have several applications: to gain insight into disease development by identifying risk factors[1], aid clinical decision making and standards,[1] and present the pathway towards truly personalised medicine[2]. Unfortunately, many published prognostic models have failed to make any clinical impact[3] and are often criticized for numerous reasons including: i) their lack of external validation (being tested on new datasets not used for model development)[4],ii) their lack of transparency (the model is a black box or the model covariates are unintuitive), iii) they are trained on small datasets and perform poorly and iv) being poorly reported[5].

One of the main factors inhibiting prognostic model validation is the lack of readily accessible data. Many researchers are limited by only having a single dataset available or similar datasets (e.g., multiple datasets collected from primary care within the same country) to train and validate a model. The Observational Health Data Sciences and Informatics (OHDSI) network is a worldwide collaboration including researchers from academia and industry. Collectively, the collaborators have a diverse set of data that spans numerous countries and continents but more importantly, the OHDSI data network enables external validation of prognostic models over a wide range of datasets in a common structure and standardized vocabulary.

In this paper, we perform a proof of concept large-scale external validation by developing models to predict myocardial infarction within 1 to 366 days of a first time prescription of celecoxib in four separate US datasets. Although this proof of concept focuses on US data, it demonstrates the potential benefits of utilizing the OHDSI frameworks to readily implement external validation. For each model, we present the model's performance when applied to the other three datasets.

**Prediction Problem**

We define the prediction problem as predicting the occurrence of myocardial infarction between 1 and 366 days after a first time prescription of celecoxib. The 'at risk' cohort is defined as people at the point of their first recorded celecoxib prescription with at least 365 days of observation time prior to the initial prescription date. The outcome cohort is defined by any inpatient recording of a diagnosis code of myocardial infarction (MI) (SNOMED concept ids 4329847 and 314666) or one of its descendants with no MI in the prior 90 days. Prediction variables included all the OHDSI FeatureExtraction package variables excluding month/year interactions and drug codes corresponding to celecoxib (>45,000 variables). The datasets used in this experiment are summarized in Table 1.

**Table 1.** The datasets used to validate and compare models.

| Database | Type | Number of people with celecoxib | Number people with MI within 1 to 366 days | Outcome percent |
|---|---|---|---|---|
| Truven CCAE | Insurance | 889,498 | 1994 | 0.22% |
| Truven Medicare | Insurance | 315,717 | 2804 | 0.89% |
| Truven Medicaid | Insurance | 34,965 | 154 | 0.44% |
| Optum | Electronic Health Records | 261,797 | 578 | 0.22% |

**Large-scale external validation framework**

To validate each model we propose publishing and sharing the model through OHDSI and calculating the discrimination of the model (AUC), the calibration of the model (calibration curve intercept and gradient) and the overall performance (brier score) on each dataset's complete data as well as the dataset's data where the test/train date are split 30:70 based on stratifying by outcome. (internal validation), see Figure 1.

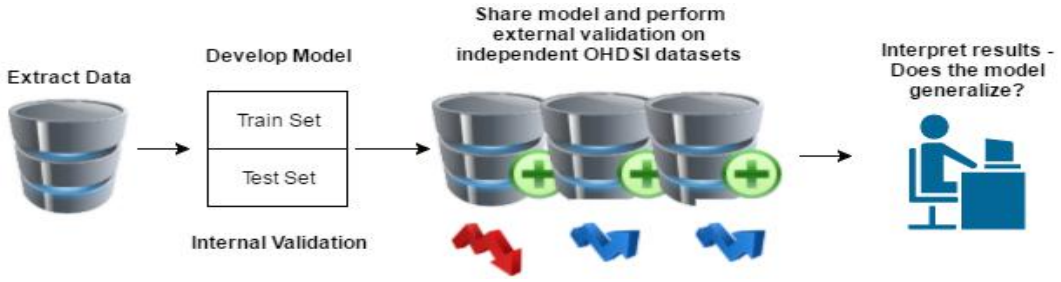


**Figure 1.** Flowchart of the process that utilizes the OHDSI network to perform large-scale independent validation.

**Results & Discussion**

The results of implementing the preliminary large-scale external validation framework for logistic regression with lasso regularization across four datasets are presented in Table 2. A different logistic regression classifier was trained on each dataset and then evaluated on the others. There was no data overlap between the four datasets.

**Table 2.** The AUC/Brier score/Calibration when validated on the three other datasets. The diagonal cells contain the test set results and the off-diagonals contain the external validation results.

| *Train dataset* | *Test/Validation dataset* | | | |
|---|---|---|---|---|
| | Truven CCAE | Truven Medicare | Truven Medicaid | Optum |
| Truven CCAE | **0.79**/0.002/(-0.000+1.010x) | **0.66**/0.009/(0.004+1.680x) | **0.79**/0.004/(0.000+0.789x) | **0.77**/0.002/(0.000+0.953x) |
| Truven Medicare | **0.73**/0.002/(-0.002+0.761x) | **0.69**/0.009/(0.001+0.905x) | **0.79**/0.004/(-0.003+0.995x) | **0.71**/0.002/(-0.001+0.705x) |
| Truven Medicaid | **0.70**/0.002/(0.000+0.911x) | **0.62**/0.009/(0.004+1.069x) | **0.81**/0.004/(-0.002+1.333x) | **0.73**/0.002/(-0.000+0.831x) |
| Optum | **0.73**/0.002/(-0.000+1.108x) | **0.63**/0.009/(0.004+0.976x) | **0.80**/0.004/(-0.001+0.960x) | **0.76**/0.002/(0.000+0.947x) |

The external validation highlighted that some models perform inconsistently across external datasets, so it is important to validate a model on as many datasets as possible to gain insight into the type of data it is suitable for. Interestingly, the models trained on other data and applied to Medicaid performed almost as well as the model developed on Medicaid and the model trained on CCAE appears to transport to the other data sets nicely. All the models performed worse on Medicare, this may be due to the Medicare dataset being very different to the other datasets (e.g., the outcome is more common and people may have more comorbidities). Identifying the cause of the drop in performance (e.g., differences between variable recordings across datasets) may help gain insight into the prediction problem and aid the development of a model that can perform well when applied to Medicare.

**Conclusion**

In this paper we presented a framework for large-scale external validation of prognostic models and presented preliminary results obtained by implementing the framework using four datasets: Truven CCAE, Truven Medicaid, Truven Medicare and Optum. The framework provided the opportunity to readily gain insight into a model's

generalizability, which is important in terms of potential clinical impact. Future work should involve expanding this study across the whole OHDSI network to validate models using data from across the world.

## References


1. Hemingway, Harry, et al. Prognosis research strategy (PROGRESS) 1: a framework for researching clinical outcomes. BMJ 2013;346: e5595.
2. Hingorani, Aroon D., et al. Prognosis research strategy (PROGRESS) 4: stratified medicine research. BMJ 2013; 346:e5793
3. Moons, Karel GM, et al. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ 2009;* 338:b606.
4. Collins, Gary S., et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC medical research methodology* 2014; 14.1:1.
5. Collins, Gary S., et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMC medicine* 2015; 13.1:1.