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

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

Aim: To develop a framework that utilizes the OHDSI network to implement large-scale model validation

BACKGROUND

Prognositic models present us with the opportunity to identify subgroups containing high risk individuals.

Many current models fail to be utlized clinically. One key factor limiting model uptake is a lack of model validation.

The OHDSI research collaboration network presents the opportunity to thoroughly evaluate prognostic models by testing their transportability on a large number of datasets across the globe.

METHOD

We performed a proof of concept large-scale external validation by developing lasso logistic regression models to predict myocardial infarction (MI) within 1 to 366 days of a first time prescription of celecoxib in four separate US datasets

Datasets:

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%

How does a model trained on one dataset perform on new datasets across the network?

RESULTS

Predicting MI - area under the ROC curve:

	Test/Validation dataset			
Train Dataset	Truven CCAE	Truven Medicare	Truven Medicaid	Optum
Truven CCAE	0.79	0.66	0.79	0.77
Truven Medicare	0.73	0.69	0.79	0.71
Truven Medicaid	0.7	0.62	0.81	0.73
Optum	0.73	0.63	0.8	0.76

Legend: Excellent: AUC 0.8+; Very Good: AUC 0.75-0.79; Good: AUC 0.70-0.74; OK: AUC 0.65-0.69; Poor: AUC 0.6-0.64

Some models (e.g. the model trained on Truven Medicare) perform inconsistently across external datasets, where as others (e.g. Truven CCAE) performed well across datasets. Interestingly, the model trained on Truven CCAE performed better on the Optum dataset compared to the model trained on Optum. The results highlight the importance of validating a model on as many datasets as possible to gain insight.

The model calibration (without recalibrating):

	Test/Validation dataset			
Train Dataset	Truven CCAE	Truven Medicare	Truven Medicaid	Optum
Truven CCAE	0.000 + 1.010x	0.004+1.680x	0.000+0.789x	0.000+0.953x
Truven Medicare	0.002 + 0.761x	0.001+0.905x	0.003+0.995x	0.001+0.705x
Truven Medicaid	0.000+0.911x	0.004+1.069x	0.002+1.333x	0.000+0.831x
Optum	0.000+1.108x	0.004+0.976x	0.001+0.960x	0.000+0.947x

CONCLUSION

Models with transportable discrimination were often not well calibrated in new data - highlighting the need to recalibrate.

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.

.

Want to be involved in continuation of study?

OHDSI
Observational Health Data Sciences and Informatics