A Prediction Model Library

Ross D. Williams, Jenna M. Reps, Peter R. Rijnbeek

Background

To facilitate the dissemination of prediction models, a centralized repository is needed that collates all models and their performance in the data network. This should also enable others to externally validate these models and update the body of evidence by uploading the results. We believe this next step in our patient-level prediction (PLP) framework is needed to have impact on clinical practice as a community. We therefore prioritized this work within the European Health Data and Evidence Network (EHDEN) project.. A reason for this is the dissemination and understanding of prediction modelling in healthcare can still be improved. The current practice leaves the different models developed by different researchers disconnected from each other. A centralized repository, or library, will collate this information together and provide improved access and usability for a range of users including regulators, clinicians and prediction researchers.

Methods

We created a relational database to store results from multiple OHDSI PLP studies. This database models the structure of the results objects generated from model development or validation. An entity relation diagram is available in Figure 1. This database can be accessed through a dedicated application which allows for the exploration of the results of multiple studies. It will also provide the ability to select models to download, this will then create a package including the required settings and cohort definitions as well as the model.
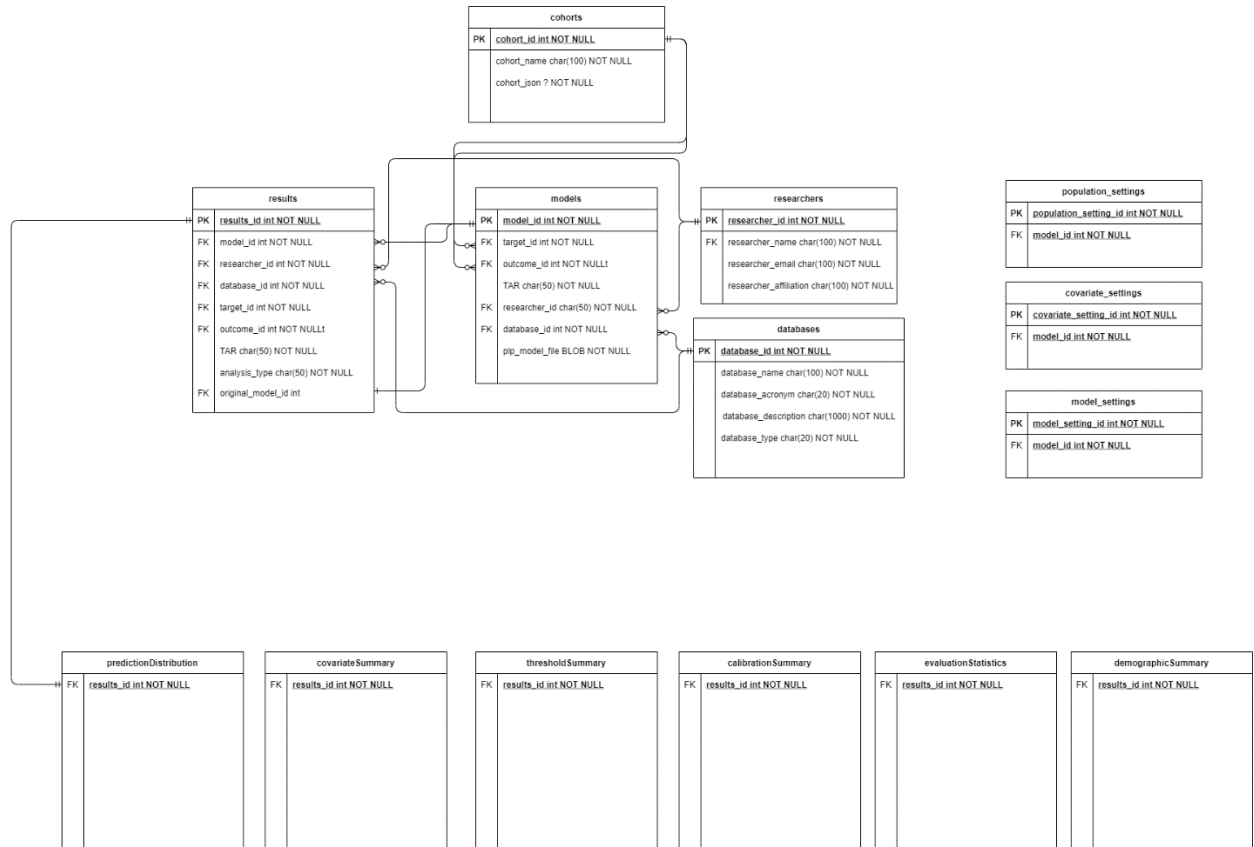
Figure 1 An entity relation diagram for the database of the prediction model library.

Results

The application contains information on more than 600 models and 20,000 different validations. An example of the application can be seen in (Figure 2).

Filters

Model:
All

Development Database
All

Validation Database
All

Target Cohort
All

Outcome Cohort
All

Time-at-risk end:
All

Column visibility | Copy | Excel | PDF   Show 10 ▼ entries   Search:

Dev	Val	T	O	Model	Covariate setting	TAR	AUC	AUPRC	T Size	O Count	Val (%)	O Incidence (%)	timeStamp
CCAE	CCAE	1 JAN 2017	AcuteMyocardialInfarction	Lasso Logistic Regression	1	(cohort start + 0) - (cohort start + 365)	0.835	0.0092	1999991	3217	25	0.1609	0
CCAE	CCAE	1 JAN 2017	Anaphylaxis	Lasso Logistic Regression	2	(cohort start + 0) - (cohort start + 365)	0.648	0.0011	1999991	999	25	0.05	0
CCAE	CCAE	1 JAN 2017	Appendicitis	Lasso Logistic Regression	3	(cohort start + 0) - (cohort start + 365)	0.573	0.0019	1999991	3105	25	0.1553	0
CCAE	CCAE	1 JAN 2017	DisIntraCoag	Lasso Logistic Regression	4	(cohort start + 0) - (cohort start + 365)	0.661	0.0002	1999991	125	25	0.0063	0
CCAE	CCAE	1 JAN 2017	Encephalomyelitis	Lasso Logistic Regression	5	(cohort start + 0) - (cohort start + 365)	0.645	0.0003	1999991	289	25	0.0145	0
CCAE	CCAE	1 JAN 2017	NonHemorrhagicStroke	Lasso Logistic Regression	6	(cohort start + 0) - (cohort start + 365)	0.751	0.0086	1999991	3820	25	0.191	0
CCAE	CCAE	1 JAN 2017	PulmonaryEmbolism	Lasso Logistic Regression	7	(cohort start + 0) - (cohort start + 365)	0.779	0.0064	1999991	2079	25	0.104	0
CCAE	CCAE	1 JAN 2017	GuillainBarreSyndrome	Lasso Logistic Regression	8	(cohort start + 0) - (cohort start + 365)	0.633	0.0001	1999991	126	25	0.0063	0
CCAE	CCAE	1 JAN 2017	HemorrhagicStroke	Lasso Logistic Regression	9	(cohort start + 0) - (cohort start + 365)	0.745	0.0017	1999991	1078	25	0.0539	0
CCAE	CCAE	1 JAN 2017	AMI_IP	Lasso Logistic Regression	10	(cohort start + 0) - (cohort start + 365)	0.833	0.0074	1999991	2181	25	0.1091	0

Showing 1 to 10 of 560 entries   Previous 1 2 3 4 5 ... 56 Next

Analysis	Dev	Val	AUC
2	CCAE	CCAE	0.648
2	IQVIA Germany	IQVIA Germany	0.56
2	JMDC	JMDC	0.629
2	MDCD	MDCD	0.553
2	MDCR	MDCR	0.542
2	Optum Claims	Optum Claims	0.583
2	Optum EHR	Optum EHR	0.56

Showing 1 to 7 of 7 entries   Previous 1 Next

Roc Plot

Calibration Plot

Figure 2 Screenshot of the library page and model validation exploration.

Conclusion

The current prototype library already provides a unique environment to interactively explore the results and evidence for prediction models developed within the OHDSI PLP framework. Work is ongoing to implement the tool as a fully functional website. This will provide an improved user experience and allow for even nicer exploration of results by our stakeholders.