Evaluating Covariate Lookback Times and the Impact on performance of Patient Level Prediction Models

Jill H. Hardin, MBA, MS, PhD[1,2], Jenna M. Reps, PhD[1,2]
[1]Observational Health Data Sciences and Informatics Community, New York, NY, USA; [2]Janssen Research and Development, Raritan, NJ, USA

Abstract

Objective: To evaluate impact of feature extraction lookback periods (14, 30, 90, 180, 365, 730, all days prior) on prediction discriminative performance for subjects with hypertensive drug exposures and development of diabetes.

Methods: We used five US observational databases to develop patient level prediction (PLP) models (IBM MarketScan® Commercial Claims and Encounters Database (CCAE), Medicare Supplemental and Coordination of Benefits Database (MDCR), Multi-State Medicaid Database (MDCD), Optum© De-Identified Clinformatics® Data Mart Database – Socio-Economic Status – (Optum), and Optum© De-Identified) Optum® de-identified Electronic Health Record Dataset (Panther). A target cohort of subjects with hypertensive drug exposures on or after 2013 and an outcome cohort of subjects with diabetes were developed. Candidate predictors from the administrative claims data that exist on or prior to the target index date were derived. Demographic, condition, drug, and procedure domains within the following lookback periods: 14, 30, 90, 180, 365, 730, and all days prior to index were evaluated. The time at risk (TAR) used was 1 day until 365 days after index. Using a bootstrapped approach, ten lasso logistic models for each lookback period were generated to create a distribution of area under the curve (AUC) metrics to evaluate the discriminative performance of the models. External validation was performed using one model and validating across four databases.

Results: The mean number of predictors increased with lookback time. The mean internally validated results show using a lookback period of 730 days is similar to using all time prior (mean AUC: 0.688 (95%CI: 0.684, 0.692) and 0.689 (95%CI: 0.682, 0.696). The AUCs using shorter lookback periods (14, 30, 90, 180, 365 days) resulted in lower AUCs (range 0.657, 0.684). The externally validated models for each lookback period had the highest AUC in CCAE, followed by MDCD, MDCR, and Panther. Externally validated results in CCAE and MDCD were more similar to the internally validated results than the externally validated results in MDCR and Panther.

Conclusions: Our results show that the lookback period that performed best for prediction models of diabetes for subjects with hypertensive drug exposure is all time prior. In future research it would be useful to see whether the results observed in this study are consistent across different prediction problems.

Research Category (please highlight or circle which category best describes your research)

Observational data management, clinical characterization, population-level estimation, patient-level prediction, other (if other, please indicate)

Background

The Observational Health Data Science and Informatics (OHDSI) collaboration have developed an end-to-end framework for developing prediction models named the patient level prediction (PLP) framework (1). The framework requires data in the OMOP common data model (CDM) (2) and enables rapid development and validation of prediction models across diverse sets of data allowing for evaluation of previously intractable patient level prediction questions. The PLP framework apply best practices for model development and evaluation, but there are still subjective choices during the model development process. An example of a subjective choice of what candidate predictors (aka features/covariates/variables) are included and the appropriate length of lookback for a given prediction question.

The PLP framework contains a library of candidate predictors that researchers can pick from when developing a model. We refer to this as the library of standard predictors. The standard predictors are constructed using data on or prior to the prediction index. The standard predictors represent several clinical domains; demographics, conditions, drugs, procedures, measurements, and observations. The standard predictors require picking specific lookback periods, for example in the 180, 365 days, or all history prior to the index date. A longer lookback time means there is more data available however it is more likely that subjects will have missing data which may confuse a model. A shorter lookback may miss capture of important features. Each lookback period gives a different perspective that may help discriminate who will experience the outcome, however, too many candidate predictors may lead to overfitting and compromise model performance. There is currently no advice available to aid researchers when deciding which candidate predictors to use (e.g., the domain type and the lookback periods).

Methods

We used four US claims and one EHR dataset to develop the prediction models and to perform external validation work. (Table 1)

Table 1 Database summary

Data Source	Coverage	Data Type	No. of Patients	%		Time, year (y)	
				Female	Male	Start	End
Optum© De-Identified Clinformatics® Data Mart Database – Socio-Economic Status – (Optum)	USA – commercially insured and medicare	**Claims** - outpatient pharmacy, inpatient and outpatient medical claims, laboratory tests for a subset of the covered lives	84,310,966	54.0	56.0	2000	2019
IBM MarketScan® Commercial Claims and Encounters Database (CCAE)	USA - commercially insured patients.	**Claims** - outpatient pharmacy, inpatient and outpatient medical claims, laboratory tests for a subset of the covered lives	152,963,555	51.2	48.8	2000	2020
IBM MarketScan® Medicare Supplemental and Coordination of Benefits Database (MDCR)	USA - medicare supplemental coverage	**Claims -** outpatient pharmacy, inpatient and outpatient medical claims, laboratory tests for a subset of the covered lives	10,087,542	55.3	44.7	2000	2019
IBM MarketScan® Multi-State Medicaid Database (MDCD)	USA – Medicaid patients	**Claims** - outpatient pharmacy, inpatient and outpatient medical claims	28,772,982	56.8	43.2	2006	2018
Optum® de-identified Electronic Health Record Dataset (Panther)	USA – commercially insured and medicare	**Electronic Health Record Dataset** - clinical information, inclusive of prescriptions as prescribed and administered, lab results, vital signs, body measurements, diagnoses, procedures, and information derived from clinical Notes using Natural Language Processing (NLP)	102,805,503	53.8	46.2	2006	2019

We created cohorts for target (new users of antihypertensive drugs after 2013 with 365 days prior observation, N=729,143(SES)) and outcome (diabetes, N=28,795 (SES)). Demographic, condition, drug, and procedure domains within the following lookback periods: 14, 30, 90, 180, 365, 730, and all days prior to index were evaluated. The time at risk (TAR) used was 1 day until 365 days after index. Using a bootstrapped approach, ten lasso logistic models for each lookback period were generated to create a distribution of area under the curve (AUC) metrics to evaluate the discriminative performance of the models. External validation was performed using one model and validating across four databases.

Results

The mean number of predictors increases with lookback time. The mean internally validated results show using a lookback period of 730 days is similar to using all time prior (mean AUC: 0.688 (95%CI: 0.684, 0.692) and 0.689 (95%CI: 0.682, 0.696). The AUCs using shorter lookback periods (14, 30, 90, 180, 365 days) resulted in lower AUCs (range 0.657, 0.684). The externally validated models for each lookback period had the highest AUC in CCAE, followed by MDCD, MDCR, and Panther. Externally validated results in CCAE and MDCD were more similar to the internally validated results than the externally validated results in MDCR and Panther.

Table 1. Internally validated AUC values from 10 replicates and external validation AUC values from application of 1 model from Optum to CCAE, MDCD, MDCR, and Panther

Lookback	Internal Validation, 10 replicates								External Validation			
Time (days)	Mean # Predictors	Mean	Max	Min	Median	Std	LCI	UCI	CCAE	MDCD	MDCR	Panther
14	693.8	0.657	0.664	0.651	0.657	0.004	0.650	0.665	0.661	0.657	0.608	0.637
30	663.7	0.659	0.663	0.654	0.660	0.003	0.653	0.665	0.664	0.661	0.618	0.637
90	753.4	0.666	0.669	0.662	0.665	0.003	0.661	0.671	0.670	0.669	0.622	0.641
180	801.4	0.674	0.679	0.670	0.673	0.003	0.668	0.681	0.677	0.663	0.627	0.636
365	852.2	0.684	0.689	0.679	0.683	0.003	0.677	0.690	0.686	0.669	0.642	0.633
730	875.9	0.688	0.692	0.684	0.688	0.002	0.684	0.692	0.688	0.664	0.646	0.634
99999	921.7	0.689	0.696	0.683	0.689	0.004	0.682	0.696	0.689	0.667	0.653	0.633

Conclusion

Our study shows that the mean number of predictors increases with more lookback time and that a lookback period of all time prior resulted in this highest AUC values. This study prompts further research to see whether all time prior is generally always sufficient as a lookback period.

References

1. Reps JM, Schuemie, M.J., Suchard, M.A., Ryan, P.B. and Rijnbeek, P.R. Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data. . J Am Med Inform Assoc 2018;25(8):969-75.

2. Voss EA MR, Matcho A, Ma Q, Knoll C, Schuemie M, et al. . Feasibility and utility of applications of the common data model to multiple, disparate observational health databases. . J Am Med Inform Assoc 2015;22(3):553-64.