Creation of a set of clinical Patient-Level Prediction benchmark tasks

Please list all authors and their affiliations in the correct order

Ross D. Williams[1], Solomon Ioannou[1], Evan Minty[2], Jenna M. Reps[1,3]

[1] Department of Medical Informatics, Erasmus MC, Rotterdam, The Netherlands, [2]O'Brien Institute for Public Health, Department of Medicine, University of Calgary, Canada, [3]Observational Health Data Analytics, Janssen Research and Development, Titusville, NJ, USA

Background

Patient-Level Prediction has the potential to transform clinical practice by allowing for highly personalized treatments for patients. To do this, we need to develop a strong body of evidence to demonstrate the performance and beneficial impact of the models. Currently models tend to not be extensively validated(1) and there exists no standard framework for methodological examination. To provide this evidence, we need to conduct robust research that allows us to contextually compare different model development and validation techniques. These analyses should be performed in similar settings to provide evidence for wider conclusions.

In this work we have created an initial set of benchmarking problems. These are relevant clinical problems that cover a range of scenarios (short and long term). Importantly, they are problem specifications that would have the potential for impact. This increases the relevance of the evidence generated about the methods research that will be conducted with them. The models will be added to the DELPHI Library (delphi.ohdsi.org) and will be available for all researchers to download as a strategus JSON object to facilitate the use within the community. The models described in this abstract are not intended for immediate clinical use.

Methods

OMOP CDM Databases

In this study we used:

- The IBM® MarketScan® Commercial Database (CCAE) includes health insurance claims across the continuum of care (e.g., inpatient, outpatient, outpatient pharmacy, carve-out behavioral healthcare) as well as enrollment data from large employers and health plans across the United States who provide private healthcare coverage for employees, their spouses, and dependents.

We conducted a study using 5 different clinical settings that have previously been used in OHDSI studies(2, 3). The three cover scores, heart failure prediction in T2DM patients, and prediction of MACE for patients undergoing major non cardiac surgery, an ongoing study.

We develop models for the five prediction tasks in one database (CCAE) using the PatientLevelPrediction framework with four different model designs:

1. Logistic regression with LASSO regularization.

Models were developed using the following the framework for PatientLevelPrediction developed by Reps et al(4). Employing a person-split where 75% of labelled data was used to learn the model with 3-fold cross validation to pick the optimal hyper-parameter and 25% of the labelled data used to internally validate the models.

Performance was assessed using the area under the receiver operating characteristic curve (AUROC).

Results

Target	Outcome	TAR	T size	O size	AUC
MNCS with measurement of creatinine	Earliest of AMI cardiac arrest or death	(cohort start + 1) - (cohort start + 1095)	15127	251	0.741
MNCS without measurement of creatinine	Earliest of AMI cardiac arrest or death	(cohort start + 1) - (cohort start + 1095)	15127	251	0.763
T2DM	Heart Failure	(cohort start + 1) - (cohort start + 1095)	530823	11735	0.747

Covid Outpatient	Hospitalisation with pneumonia	(cohort start + 1) - (cohort start + 1095)	1441291	23921	0.825
Covid Outpatient	Hospitalisation with pneumonia and intensive services	(cohort start + 1) - (cohort start + 1095)	1455989	4451	0.837
Covid Outpatient	Fatality	(cohort start + 1) - (cohort start + 1095)	1459228	468	0.566

Conclusion

We produced models with a range of cohort sizes (15127-1459228) across 3 disease domains. The models ranged in performance form (0.566-0.837). The models will be added to the DELPHI library and are considered a starting point for the benchmarking project within the PLP framework.

References

1. Yang C, Kors JA, Ioannou S, John LH, Markus AF, Rekkas A, et al. Trends in the conduct and reporting of clinical prediction model development and validation: a systematic review. J Am Med Inform Assoc. 2022;29(5):983-9.
2. Williams RD, Markus AF, Yang C, Duarte-Salles T, DuVall SL, Falconer T, et al. Seek COVER: using a disease proxy to rapidly develop and validate a personalized risk calculator for COVID-19 outcomes in an international network. BMC Med Res Methodol. 2022;22(1):35.
3. Williams RD, Reps JM, Kors JA, Ryan PB, Steyerberg E, Verhamme KM, et al. Using Iterative Pairwise External Validation to Contextualize Prediction Model Performance: A Use Case Predicting 1-Year Heart Failure Risk in Patients with Diabetes Across Five Data Sources. Drug Saf. 2022;45(5):563-70.
4. Reps JM, Williams RD, You SC, Falconer T, Minty E, Callahan A, et al. Feasibility and evaluation of a large-scale external validation approach for patient-level prediction in an international data network: validation of models predicting stroke in female patients newly diagnosed with atrial fibrillation. BMC Med Res Methodol. 2020;20(1):102.