PatientLevelPrediction WG

2024 Updates

What is the PatientLevelPrediction WG?

We are a group of researchers interested in best practices for developing and implementing healthcare prediction models using observational data.

- We collaborate to answer method's research questions
- We run network studies (e.g., validating existing models across OHDSI databases)
- We develop prediction models (e.g., COVID prediction models)
- We maintain an R package to aid methods research/model development/network studies
- We discuss ideas about prediction model development

Example projects

- External validation performance estimation using summary statistics (**Chen Yanover**)
- Comparison of deep learning and conventional strategies for disease onset prediction (**Henrik John**)
- Insights from a prognostic model implementation review (**Alexander Saelmans**)
- Seek Cover validation over time (**Egill Fridgeirsson**)
- Developing a set of benchmark tasks (**Ross Williams**)

Chen Yanover
KI Research Institute

Henrik John
Erasmus MC

Alexander Saelmans
Erasmus MC

Egill Friðgeirsson
Erasmus MC

Ross Williams
Erasmus MC

More examples

- **Federated learning** research – how can we do federated learning
- Investigating the impact of the **metric used to select hyper-parameters**
- How does **data type/density** impact prediction
- **Counterfactual** prediction methods
- …

- **Do you have any prediction research topics?**

How to get involved?

- Sign up to the workgroup via https://forms.office.com/Pages/ResponsePage.aspx?id=lAAPoyCRq0q6TOVQkCOy1ZyG6Ud_r2tKuS0HcGnqiQZUOVJFUzBFWE1aSVlLN0ozR01MUVQ4T0RGNyQlQCN0PWcu (make sure to tick 'PLP: Patient-Level Prediction')

- Reach out to Ross Williams or myself (Jenna Reps)

- We meet every second Wednesday of the month @ 9am ET (next meeting is August 14th)

R Package: https://github.com/OHDSI/PatientLevelPrediction

If you want to easily develop models using your OMOP CDM data try the package: remotes::install_github('OHDSI/PatientLevelPrediction')

or

If you want to contribute to an R package that makes model development/validation a smooth process: let us know and start coding.

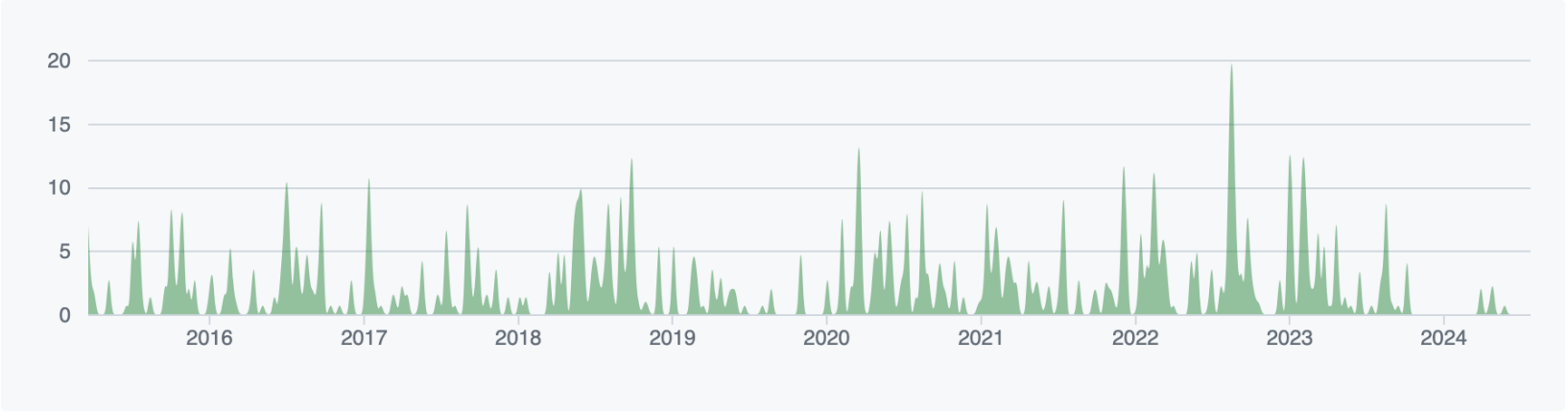


Figure 1. This plot shows commits by date to the GitHub repository for PatientLevelPrediction

Estimating Model Performance in External Databases from Limited Statistical Characteristics: A Benchmark Study

Tal El-Hay, Jenna M Reps, Chen Yanover

Motivation

- Performance of prediction models may deteriorate when applied to data sources not used for training

→ External validation

- Challenging, potentially iterative, task

→ Estimate performance in an external data source from its limited descriptive statistics*

  – Seek patient-level weights that induce internal weighted statistics that are “similar” to the external ones

  – Compute performance metrics using the internal weighted cohort

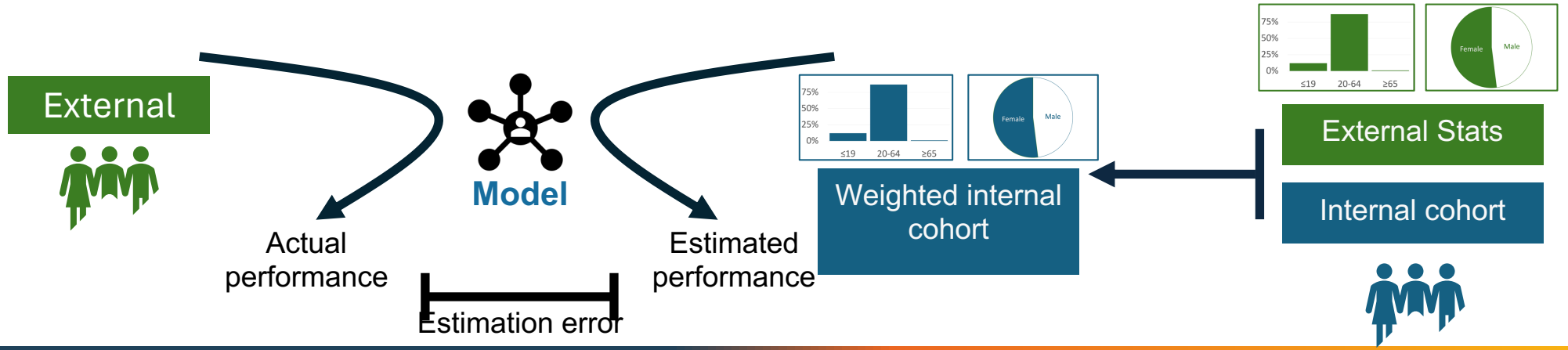
  * Extracted per-study or literature-based

Benchmark

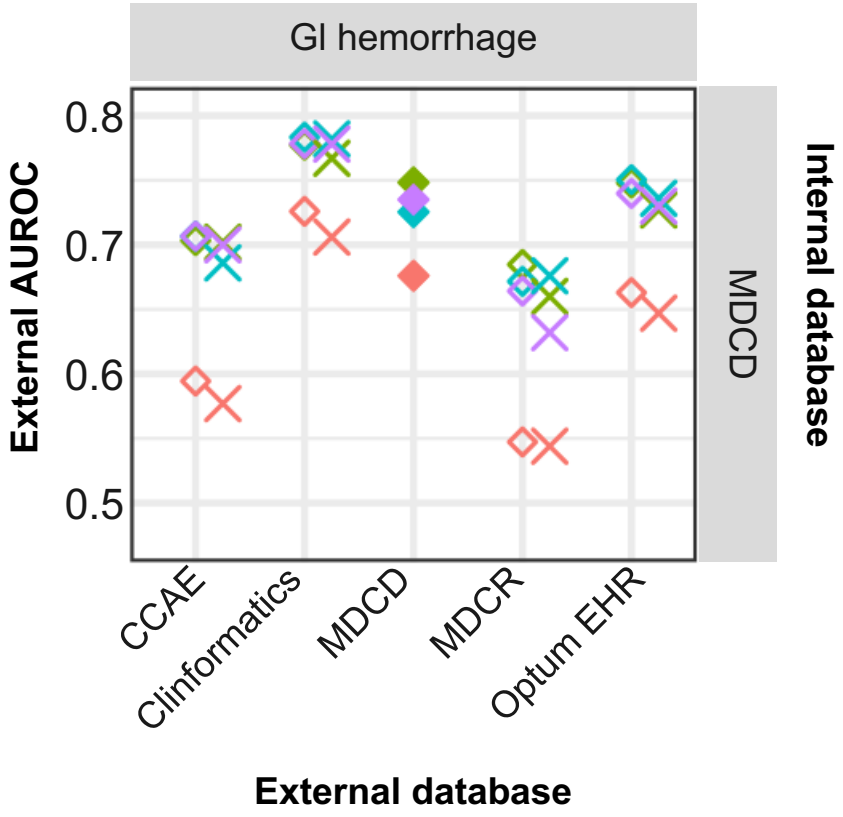
CCAE	MDCD	MDCR	Optum EHR	Clinformatics
Internal				
External				
External				
External				
External				Internal

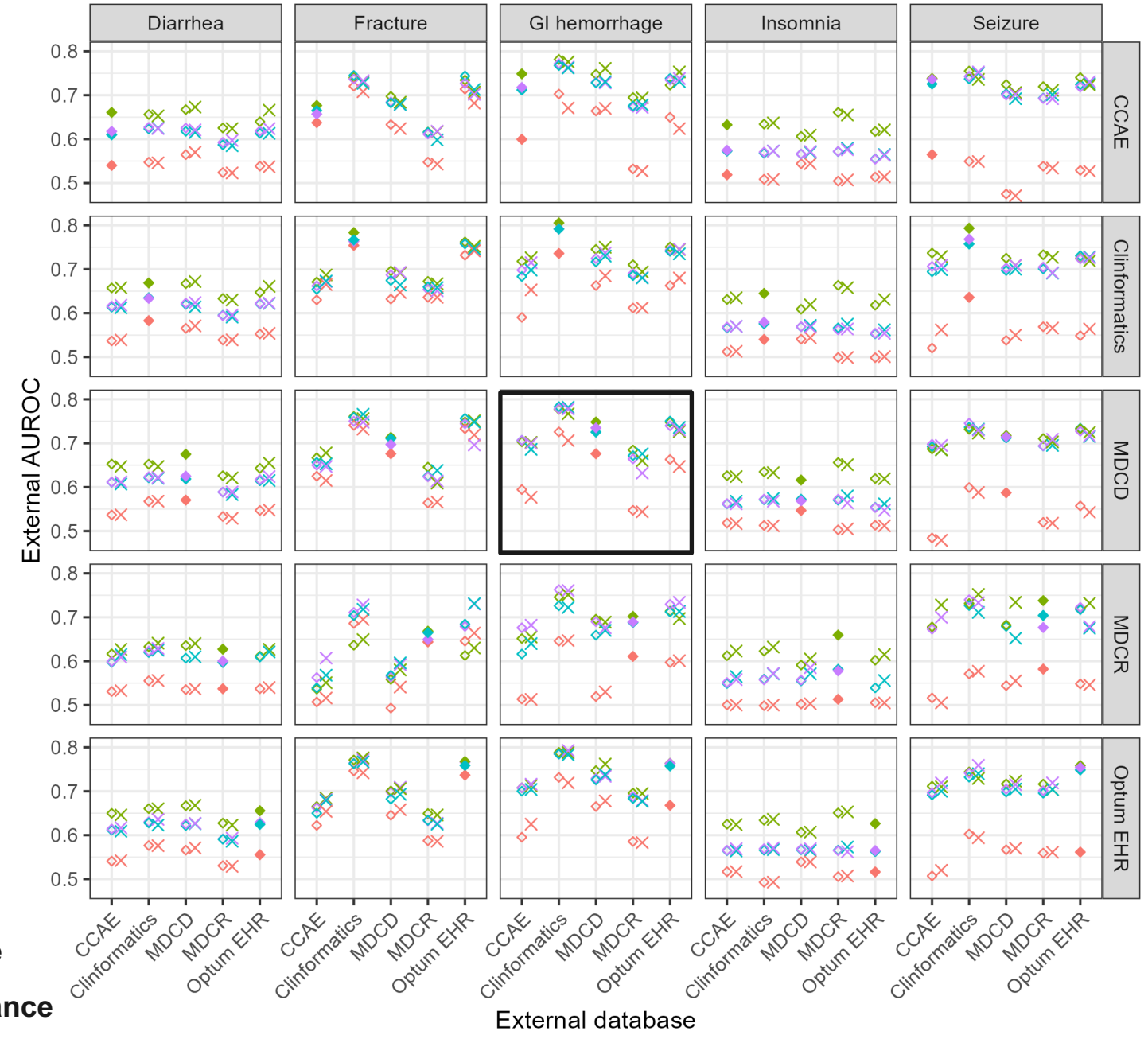
× 5 outcomes

× 4 model variants



Results





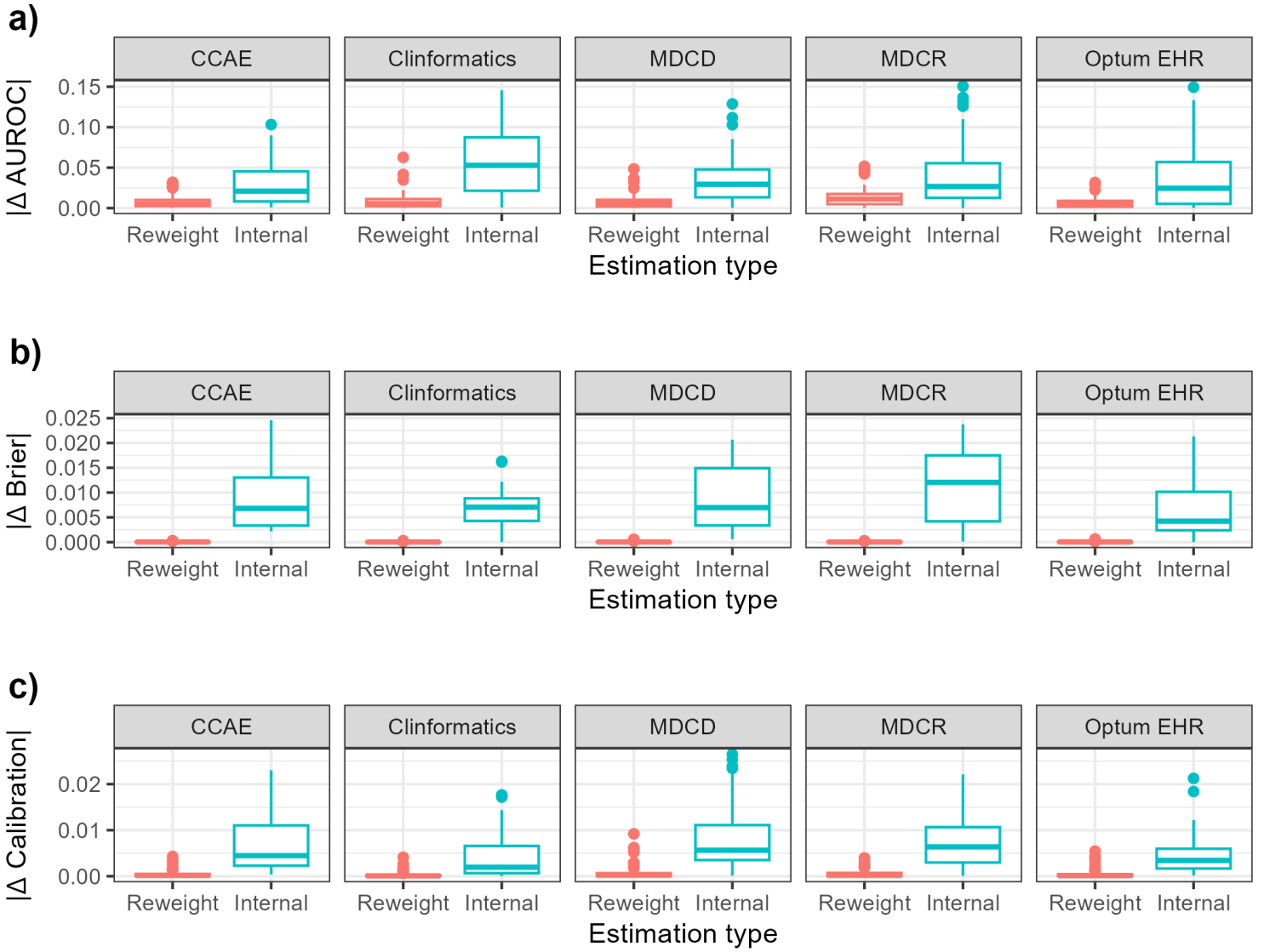
● LR, age-sex ● LR, phenotypes ◇ Actual performance

● LR, full ● LR, full x Estimated performance

Results

Accurate estimations, typically:

Discrimination (AUROC)	≤ 0.02
Overall accuracy (Brier score)	≤ 0.001
Calibration (calibration-in-the-large)	≤ 0.005



Conclusions

- Accurate estimations for all metrics, but depending on
  - Diversity of internal cohort re to external cohort (one-sided positivity)
  - Proper selection of features
- Some recommendations, diagnostics
- Useful in pre-deployment assessment, expedite collaborative model development

El-Hay & Yanover, *Proceedings of the Conference on Health, Inference, and Learning (CHIL)*, 2022

Algorithm's GitHub repository

Comparison of deep learning and conventional strategies for disease onset prediction

- OHDSI Network Study
- Collaborators:

Henrik John[1], Chungsoo Kim[2], Jan Kors[1], Junhyuk Chang[3], Hannah Morgan-Cooper[4], Priya Desai[4], Chao Pang[5], Peter Rijnbeek[1], Jenna Reps[1,6], Egill Fridgeirsson[1]

[1]Erasmus University Medical Center, Rotterdam, The Netherlands

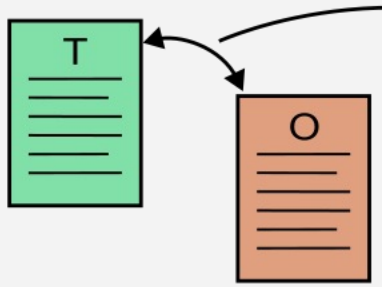
[2]Yale School of Medicine, New Haven, CT, United States

[3]Ajou University Graduate School of Medicine, Republic of Korea

[4]Stanford School of Medicine and Stanford Health Care, Palo Alto, CA, United States

[5]Columbia University Irving Medical Center, New York, NY, United States

[6]Janssen Research and Development, Titusville, NJ, United States



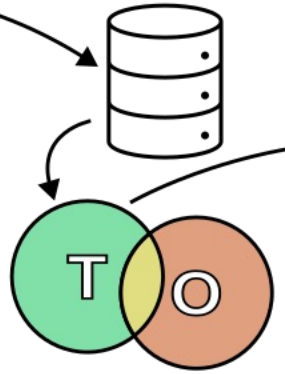
Step 1
Prediction problem

Definition of target-outcome pairs for onset prediction.

T Target cohort
O Outcome cohort

Prediction problems

- Dementia in persons above the age of 55.
- Bipolar disorder in persons newly diagnosed with major depressive disorder
- Lung cancer in persons who are cancer-free at first outpatient visit in the year.

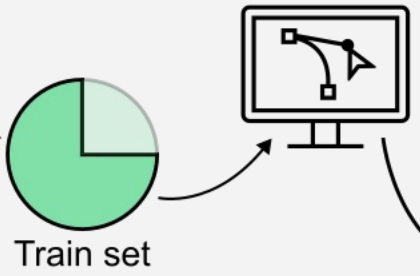


Step 2
Database extraction

Extract target and outcome cohort from database. Label intersection of cohorts as persons with the outcome in the target.

Databases

- IPCI
- AUSOM
- OPEHR
- OPSES
- STARR-OMOP
- CUIMC

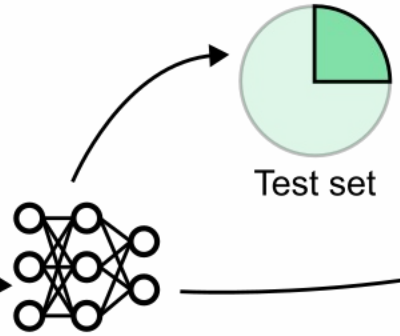


Step 3
Model development

Partition data into train and test set. Develop models for various prediction methods on train set.

Prediction methods

- Logistic regression
- Xgboost
- ResNet
- Transformer

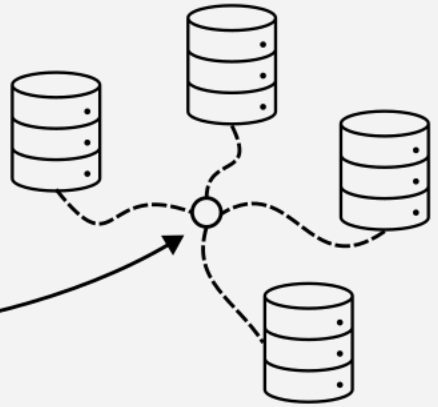


Step 4
Internal validation

Evaluate discrimination and calibration performance of models on test set.

Evaluation metrics

- Discrimination: Area under the receiver operating characteristic curve
- Calibration: Eavg or Integrated Calibration Index

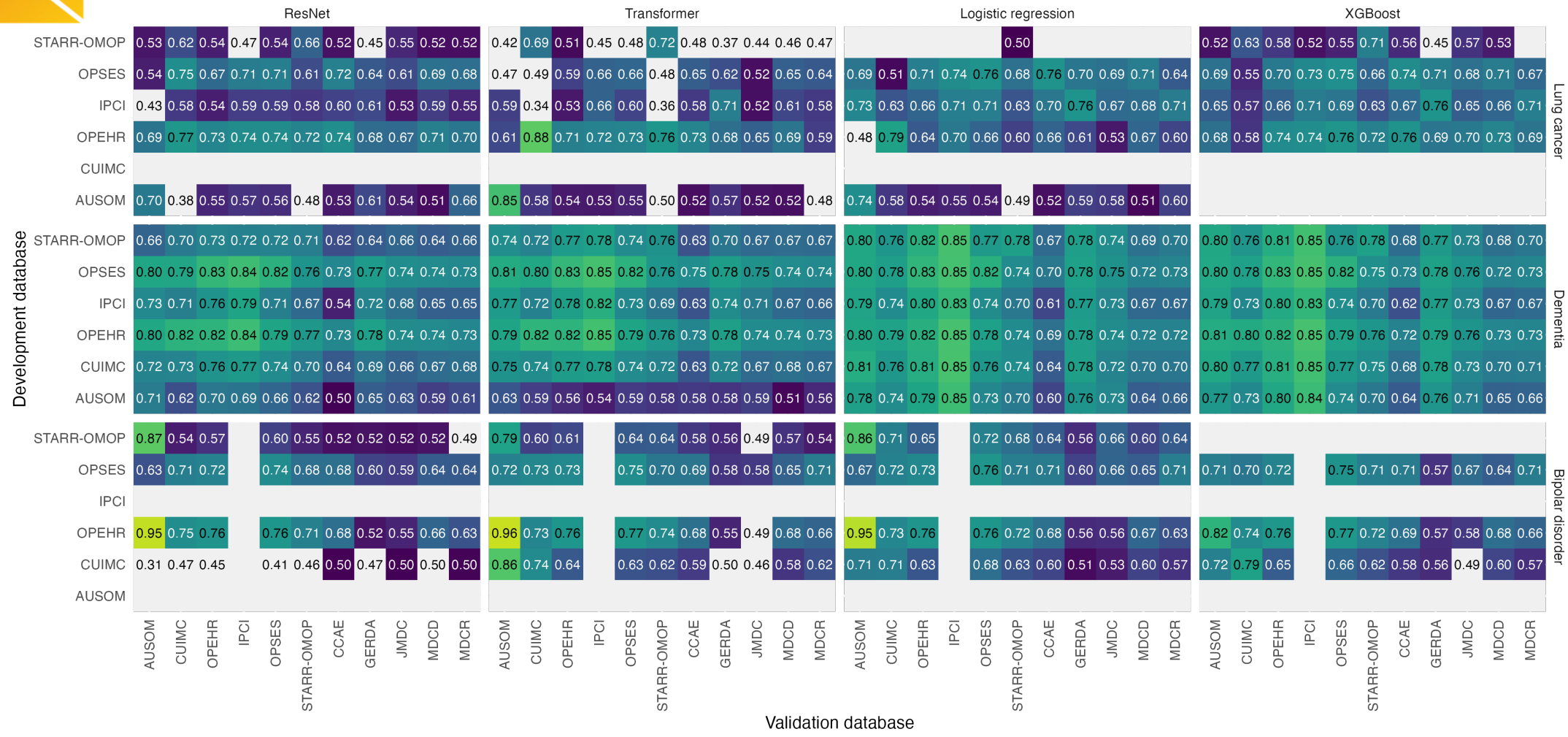


Step 5
External validation

Evaluate discrimination and calibration performance of models on external data sources.

Databases

- IPCI
- AUSOM
- OPEHR
- OPSES
- STARR-OMOP
- CUIMC
- GERDA
- JMDC
- CCAE
- MDCD
- MDCR


ResNet
Transformer
Logistic regression
XGBoost
Lung cancer
Dementia
Bipolar disorder
Development database
STARR-OMOP
OPSES
IPCI
OPEHR
CUIMC
AUSOM
Validation database
AUSOM
CUIMC
OPEHR
IPCI
OPSES
STARR-OMOP
CCAE
GERDA
JMDC
MDCD
MDCR
AUROC
0.5
0.6
0.7
0.8
0.9
1.0

Ranking of prediction method based on discrimination performance (AUROC) for:

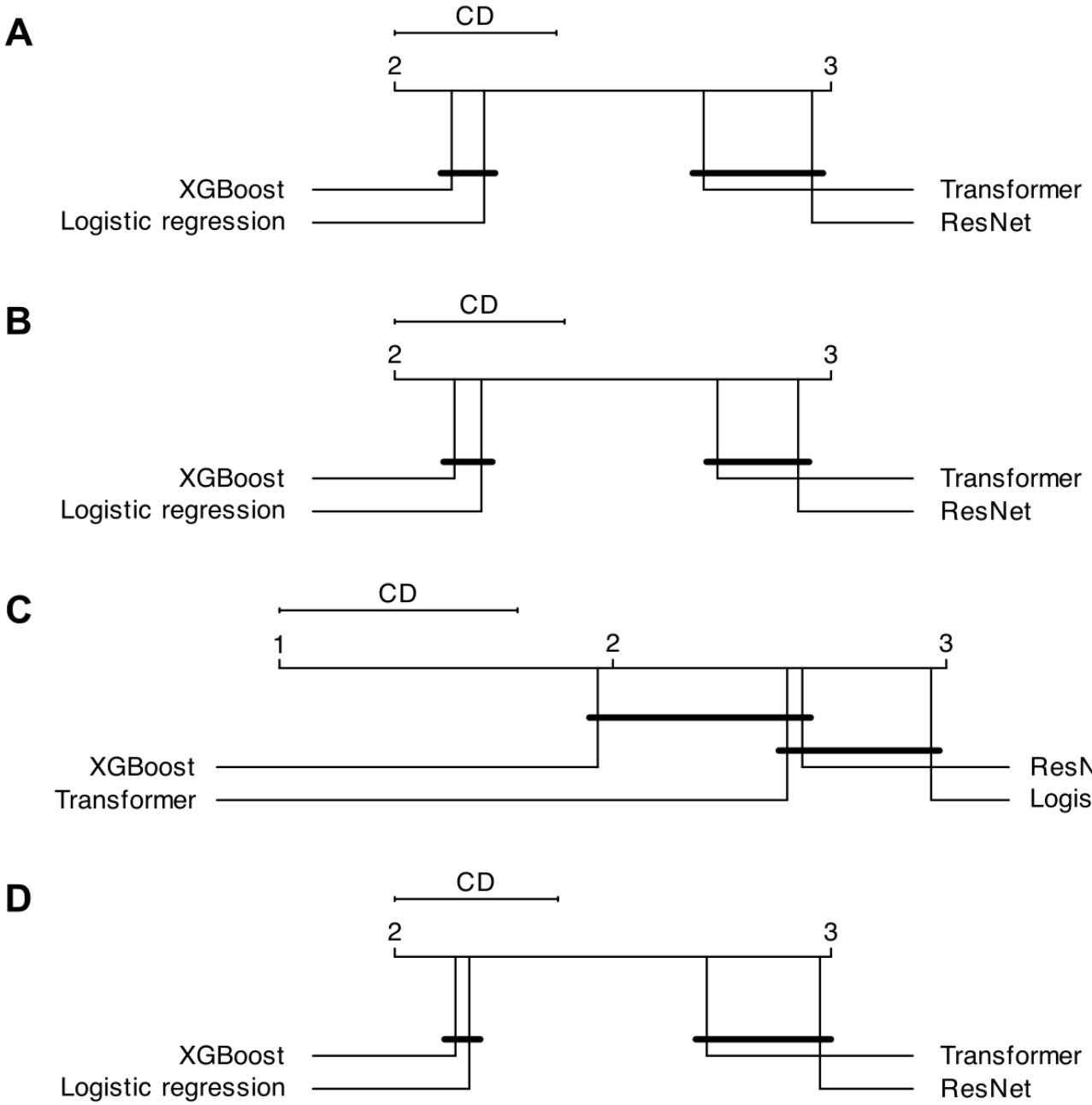
A internally and externally validated models

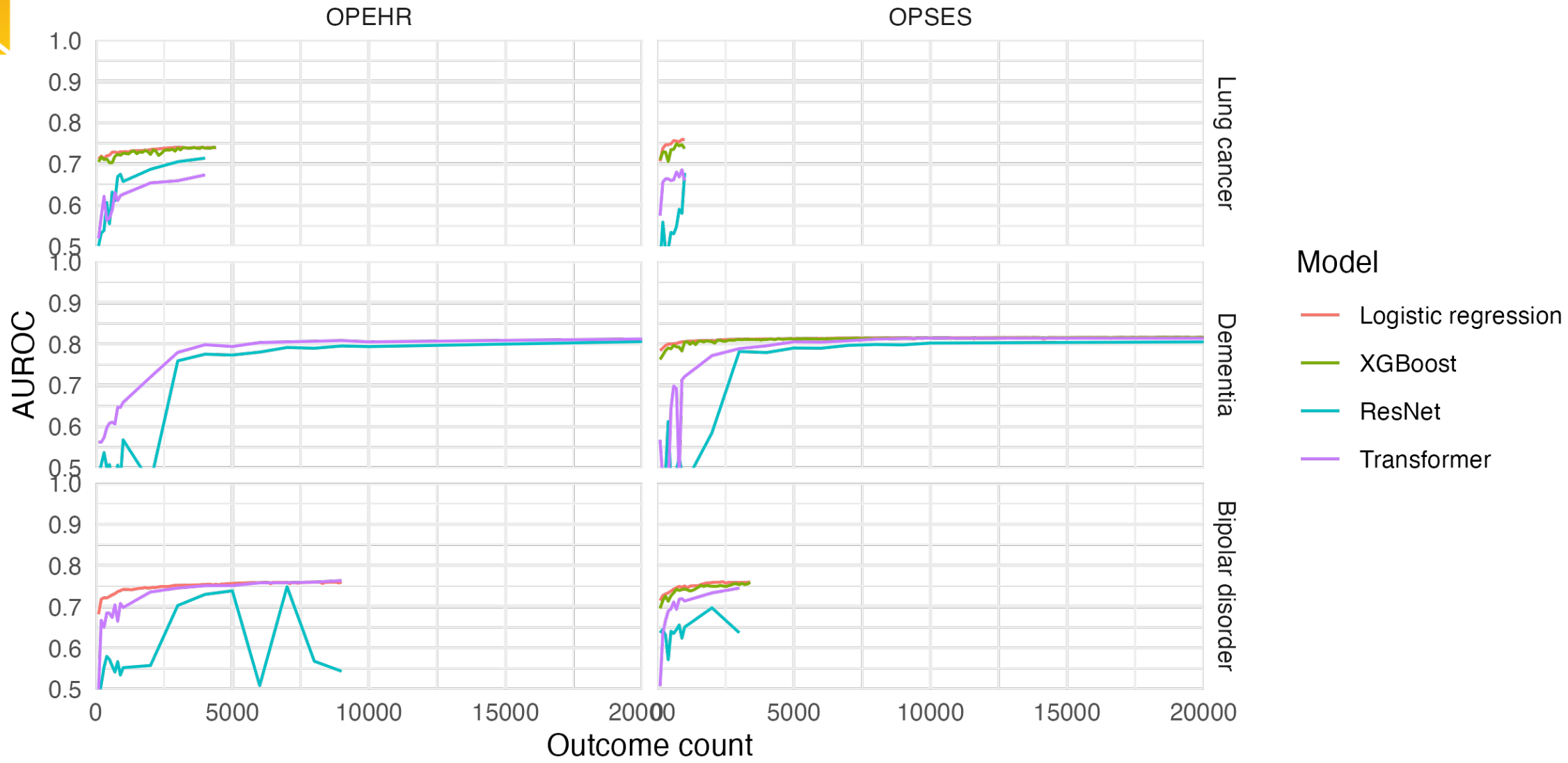
No significant difference for internally validated models

B externally validated models

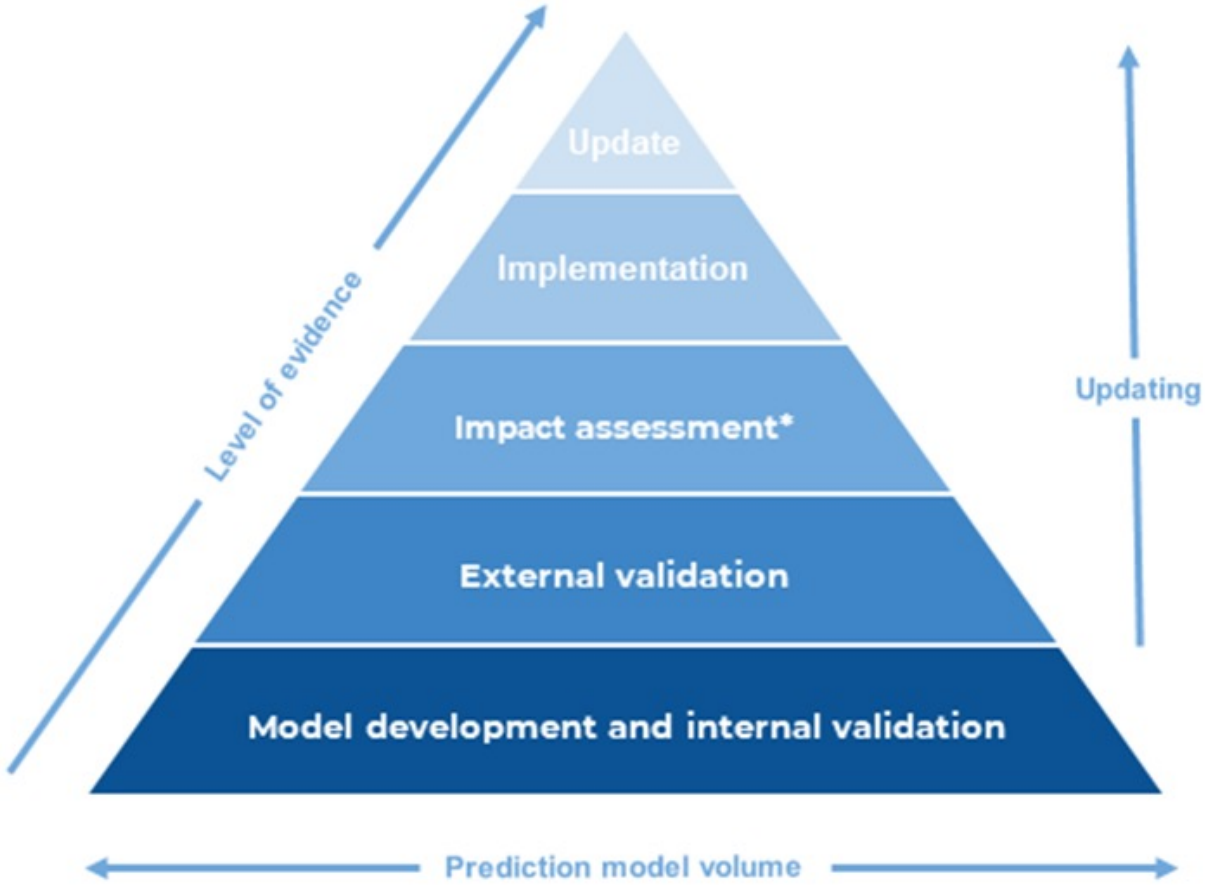
C models developed and validated on OPSES and OPEHR

D models developed on IPCI, AUSOM, STARR-OMOP, and CUIMC and validated across all data sources except OPSES and OEHR

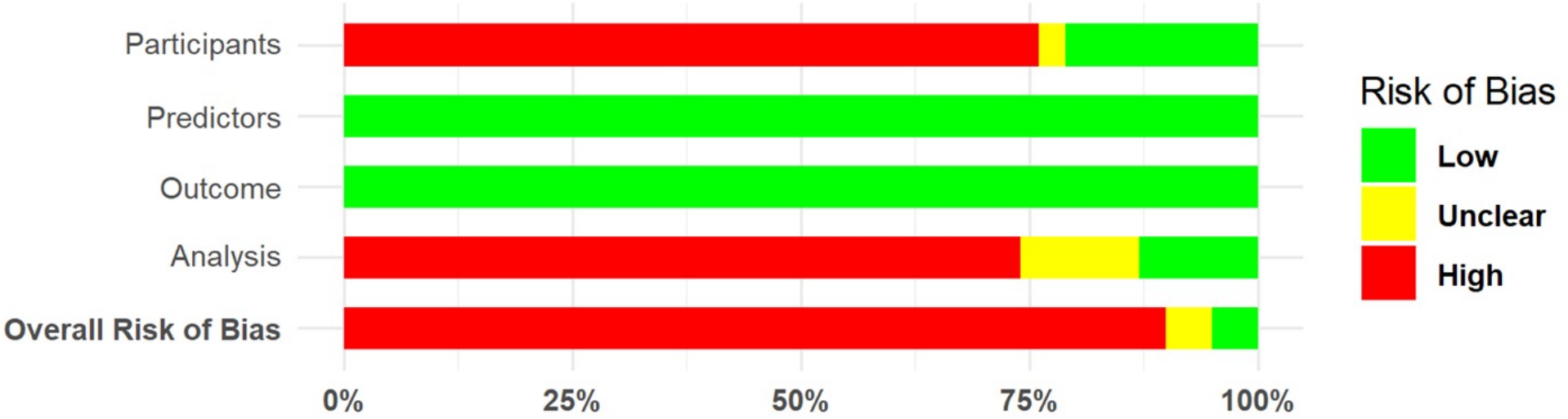



OPEHR
OPSES
Lung cancer
Dementia
Bipolar disorder
AUROC
1.0
0.9
0.8
0.7
0.6
0.5
0
5000
10000
15000
20000
Outcome count
Model
Logistic regression
XGBoost
ResNet
Transformer

Implementation and Updating of Clinical Prediction Models
A Systematic Review (Alexander)

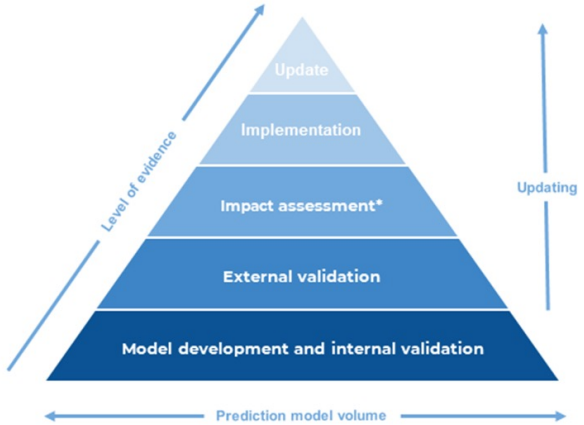


Implementation and Updating of Clinical Prediction Models
A Systematic Review



Implementation and Updating of Clinical Prediction Models
A Systematic Review

- AI multidisciplinary teams
- Prospective registry
- Research on updating strategy best practice



Covid validation over time

- In beginning of the Covid Pandemic OHDSI developed a prediction model
- Now after the pandemic is over we want to investigate performance of such models during the pandemic

Williams *et al.*
BMC Medical Research Methodology (2022) 22:35
https://doi.org/10.1186/s12874-022-01505-z

BMC Medical Research Methodology

RESEARCH Open Access

Seek COVER: using a disease proxy to rapidly develop and validate a personalized risk calculator for COVID-19 outcomes in an international network

Ross D. Williams[1†], Aniek F. Markus[1†], Cynthia Yang[1], Talita Duarte-Salles[2], Scott L. DuVall[3], Thomas Falconer[4], Jitendra Jonnagaddala[5], Chungsoo Kim[6], Yeunsook Rho[7], Andrew E. Williams[8], Amanda Alberga Machado[9], Min Ho An[10], María Aragón[2], Carlos Areia[11], Edward Burn[2,12], Young Hwa Choi[13], Iannis Drakos[14], Maria Tereza Fernandes Abrahão[15], Sergio Fernández-Bertolín[2], George Hripcsak[4], Benjamin Skov Kaas-Hansen[16,17], Prasanna L. Kandukuri[18], Jan A. Kors[1], Kristin Kostka[19], Siaw-Teng Liaw[5], Kristine E. Lynch[3], Gerardo Machnicki[20], Michael E. Matheny[21,22], Daniel Morales[23], Fredrik Nyberg[24], Rae Woong Park[25], Albert Prats-Uribe[12], Nicole Pratt[26], Gowtham Rao[27], Christian G. Reich[19], Marcela Rivera[28], Tom Seinen[1], Azza Shoaibi[27], Matthew E. Spotnitz[4], Ewout W. Steyerberg[29,30], Marc A. Suchard[31], Seng Chan You[25], Lin Zhang[32,33], Lili Zhou[18], Patrick B. Ryan[27], Daniel Prieto-Alhambra[12], Jenna M. Reps[27†] and Peter R. Rijnbeek[1*†]

Williams, R.D., Markus, A.F., Yang, C. 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 22, 35 (2022). https://doi.org/10.1186/s12874-022-01505-z

Covid validation over time - motivation

- validate the model every x months
  - Depends on outcome count how frequently we can validate
- How is performance affected by different events
  - Start of vaccinations
  - Different strains
- How do strategies such as using a disease proxy or waiting for enough data compare in performance down the line
- How is this affected by different model updating strategies

Covid validation over time

- The study repo:
  - https://github.com/ohdsi-studies/PandemicPrediction
- Are in the study design phase
  - If you want to get involved right away please be in touch:
  - e.fridgeirsson@erasmusmc.nl or study repo issue tracker
- Want to run as a network study
  - Will advertise that later and reach out to original data partners

Benchmark tasks

Why?

It is important to have a consistent set of prediction tasks and metrics to enable fair comparison of methods

Requirements:

- Must be clinically useful
- Must be diverse (topic, size, …)
- Must be feasible in observational data

Benchmarks – What we did:

We crowd sourced prediction tasks:

- We asked researchers in OHDSI workgroups that are clinical to provide us prediction tasks they would find useful.
- We asked OHDSI EU symposium prediction workgroup attendees to suggest prediction tasks they would find useful.
- We identified existing models that have been published

https://ohdsi.github.io/PatientLevelPrediction/articles/BenchmarkTasks.html

We are currently creating the phenotypes for each task and creating an R package that will create the datasets for the benchmark tasks.

Benchmarks – Next steps

Next Steps:

- We want to evaluate the performance of standard models across a network of databases for the benchmark tasks
- We will write a journal paper (everyone who contributed to the tasks will be a coauthor)

Importance:

Now methods research can use these clinically meaningful benchmark tasks

Questions

Join the OHDSI prediction journey!

Thank you for listening!

We hope to see some new faces in our next monthly call.

Email: jreps@its.jnj.com