On the Patient Level Prediction (PLP) Working Group in OHDSI

Md. Shamsuzzoha Bayzid

Slides are adapted from the one presented in PLP meeting

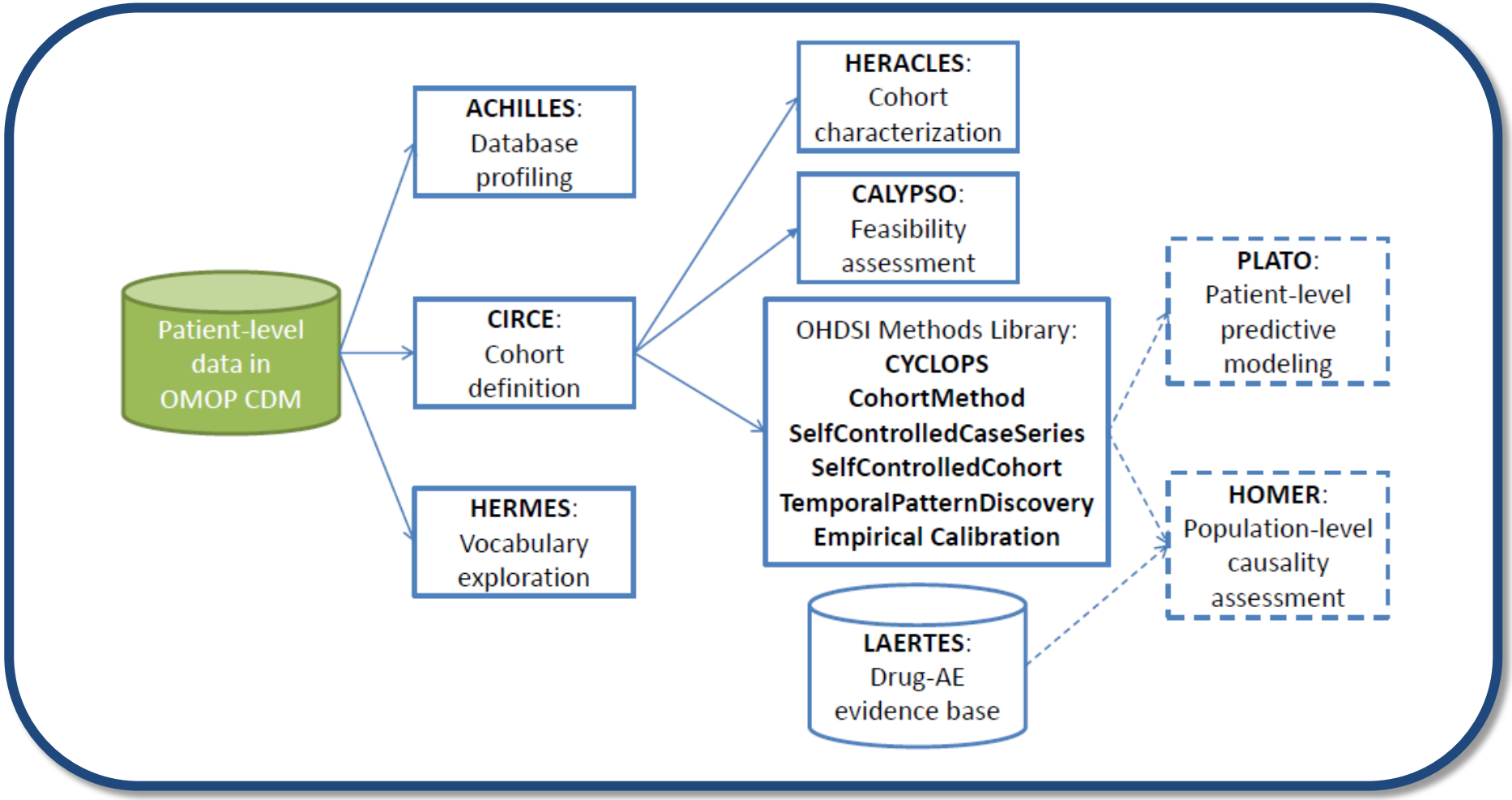
University of Texas at Austin

OHDSI

- The **O**bservational **H**ealth **D**ata **S**ciences and **I**nformatics (**OHDSI**) program is a multi-stakeholder, interdisciplinary collaborative to create open source solutions that bring out the value of observational health data through Large scale analytics.

- OHDSI has established an international network of researchers and observational health databases with a central coordinating center housed at Columbia University

Software tools



Patient Level Prediction Work-Group

The overall objective of this WG is to establish a standardized process for developing accurate and well-calibrated patient-centered predictive models that can be utilized for multiple outcomes of interest and can be applied to observational healthcare data from *any* patient subpopulation of interest

Lead:	Peter Rijnbeek	p.rijnbeek@erasmusmc.nl
Co-lead:	Jenna Reps	Jreps@ITS.JNJ.com

Clinicians are confronted with prediction questions on a daily basis. What options do they have?

Deny ability to predict at the individual patient level

Quote an overall average to all patients

Utilize knowledge and personal experience

Provide a personalized prediction based on an advance clinical model

Building a clinical prediction model is not an easy task

- Often prediction models proved to be not generalizible to other study populatons
  -> external model validation is crucial

- Should be easy to incorporate in clinical practice
  -> not too complex comprehensible models are preferred

- Focus is often on accurary and clinical usefullness is ignored
  -> ideally, the model is used in a trial with health as outcome

- Often prediction models are reported inadequately -> Transparent Reporting of a multivariable prediction Model for Individual Prognoss Or Diagnosis (TRIPOP: www.tripod-statement.org)

Other application areas

- Patient-level predictive modeling could be applied to the prediction of side effects and could be part of new risk minimization strategies, e.g. which patients are more likely to get bleeds after Warfarin use so better monitoring could be proposed?

- Ascertainment of patient characteristics that result in a beneficial outcome of the treatment

- Selection and adjustment in design and analysis of RCTs

Interestingly..

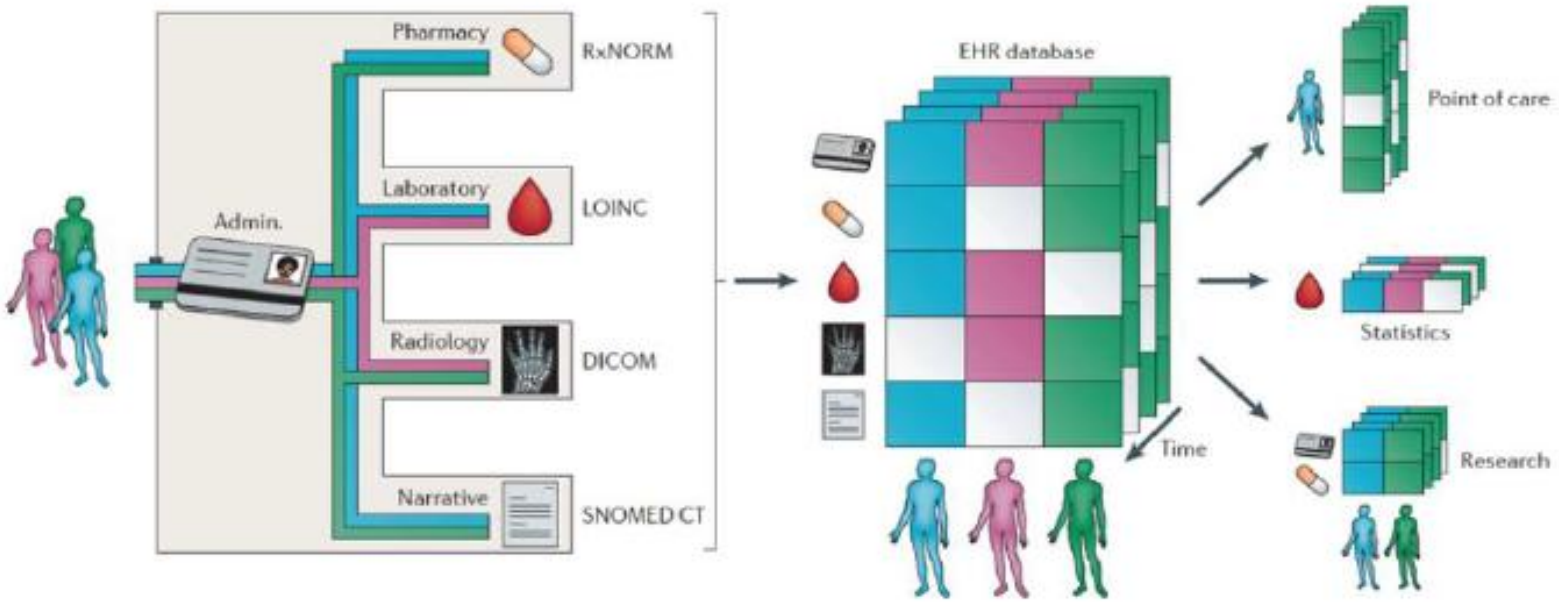
Most clinically used models are estimated using small datasets and contain a limited set of patient characteristics.

For example, in a review of 102 prognostic models in traumatic brain injury showed that three quarters of the models were based on samples with less than 500 patients [5].

- Low number forces researcher to make model assumptions
- Covariate selection is often strongly guided by expert knowledge

Massive-scale, patient-specific predictive modeling can become reality

- Large amount of EHR data available
- Initiatives like OHDSI and the OMOP CDM are instrumental



Research objectives

Objective 1: Develop a systematic process to estimate patient-level prediction models using real-world EHR data represented as observational data in the OMOP common data model

Objective 2: Develop an evaluation framework to assess the validity of the estimated models and provide insight in the effect of the design choices made

Person-Level Assessment of Treatment Outcomes (PLATO) tool integrated OHDSI framework

PLP Package

An R package for building patient level predictive models using data in
OMOP Common Data Model format.

PLP Package

- Takes a cohort and outcome of interest as input
- Extract necessary data from the database in OMOP model
- Use a large set of covariates including for example all drugs, diagnoses, procedures, as well as age, comorbidity, etc.
- Large scale regularized regression (CYCLOPS package) to fit the predictive models
- Includes functions for evaluating the predictive models
- Supported outcome models are: logistic, survival etc.

PLP Extra Methods

A general framework to enable a range of feature engineering, feature selection and classification models to be included.

Framework

- Extract plp data
- Censor/filter plp data – by date, observation period, age, gender (can filter classes differently)
- Describe plp data – visualisation of the data (e.g., prevalence by year)
- **Develop model** (combines multiple steps…)
- **Compare models**

Currently…

- Classifiers included:
  - Lasso logistic regression (plp default)
  - Neural network (caret)
  - Naïve Bayes (caret)
  - Gradient boosting machines (h2o)
  - Random forest (h2o)
  - Elastic net logistic regression (h2o)
  - Coming soon: stacking ensemble (h2o ensemble)
- Other suggestions?

R package

- Preliminary package at github PatientLevelPrediction branch extraMethods

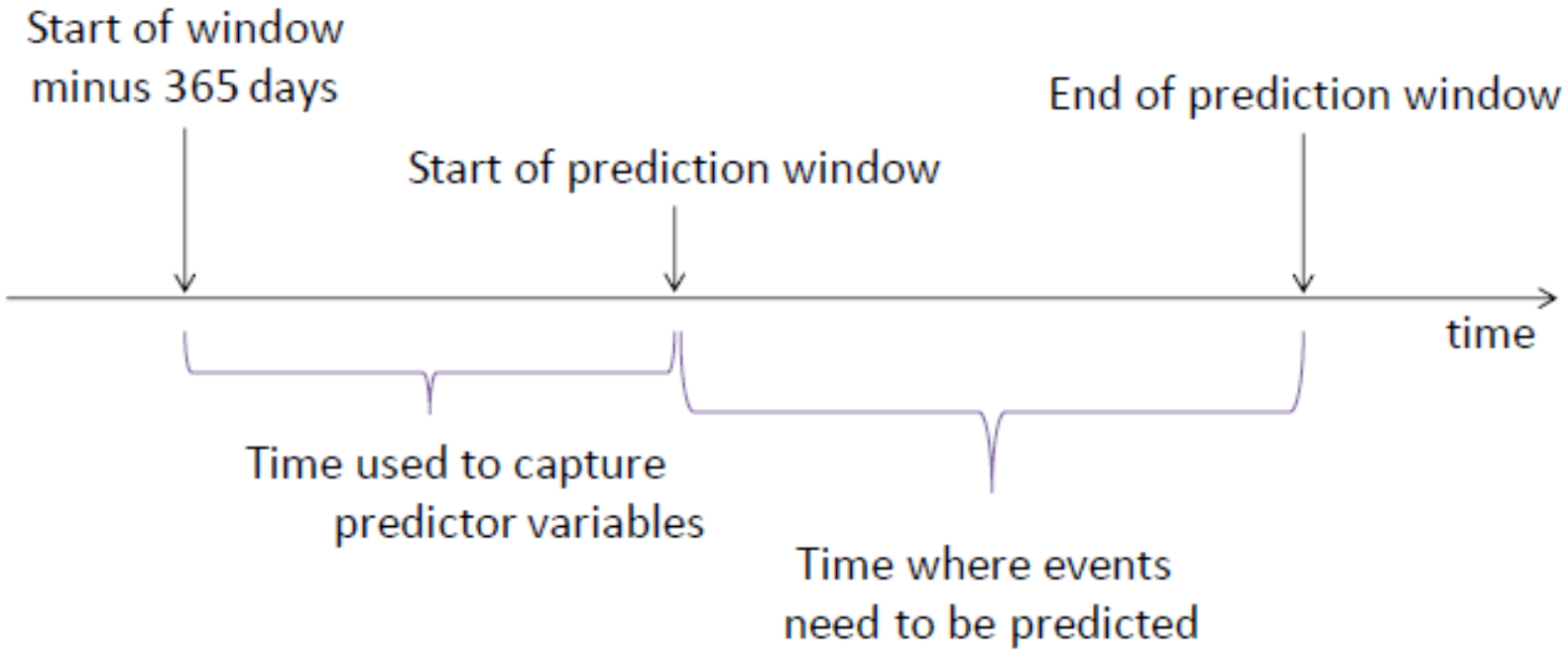
- https://github.com/OHDSI/PatientLevelPrediction/tree/extraMethods

Thank You

PLP Package

Given a cohort of interest and an outcome of interest, the package can use data in the Common Data Model to build a large set of features. These features can then be used by the Cyclops package to fit a predictive model. Also included are function for evaluating the predictive models

Defining the prediction window



Patient-Level Prediction Workgroup

Objective:

The overall objective of this WG is to establish a standardized process for developing accurate and well-calibrated patient-centered predictive models that can be utilized for multiple outcomes of interest and can be applied to observational healthcare data from any patient subpopulation of interest.

Lead:	Peter Rijnbeek	p.rijnbeek@erasmusmc.nl
Co-lead:	Jenna Reps	jreps@ITS.JNJ.com

Wiki page created on the OHDSI website: link

Github: https://github.com/OHDSI/PatientLevelPrediction

How to find the hay in the haystack?

- By nature sparse and irregular-spaced data
- Clinical narratives to be mined
- Dimensionality reduction methods need to be assessed and developed to deal with the massive amount of data
- We need to use the valuable temporal information
- How to incorporate expert knowledge

- Proper performance measures needed

OHDSI

- **Clinical characterization**
  - Natural History: Who are the patients who have diabetes? Among those patients, who takes metformin?
  - Quality improvement: what proportion of patients with diabetes experience disease-related complications?
- **Population-level estimation**
  - Safety surveillance: Does metformin cause lactic acidosis?
  - Comparative effectiveness: Does metformin cause lactic acidosis more than glyburide?
- **Patient-level prediction**
  - Personalized medical care: Given everything you know about me and my medical history, if I start taking metformin, what is the chance that I am going to have lactic acidosis in the next year?

Clinical prediction models examples

Many prediction models have been developed in patient subgroups at higher risk that need more intensive monitoring:

- Prediction of 30-day mortality after an acute myocardial

Models have been developed for asymptomatic subjects in the population:

- Framingham risk functions for cardiovascular disease, or models for breast cancer screening.

HERACLES: Quality of care

» Health Enterprise Resource and Care Learning Exploration System (HERACLES) - descriptive statistics about a cohort

- cohort summarization and visualization tool