Complementary evidence to inform the patient journey
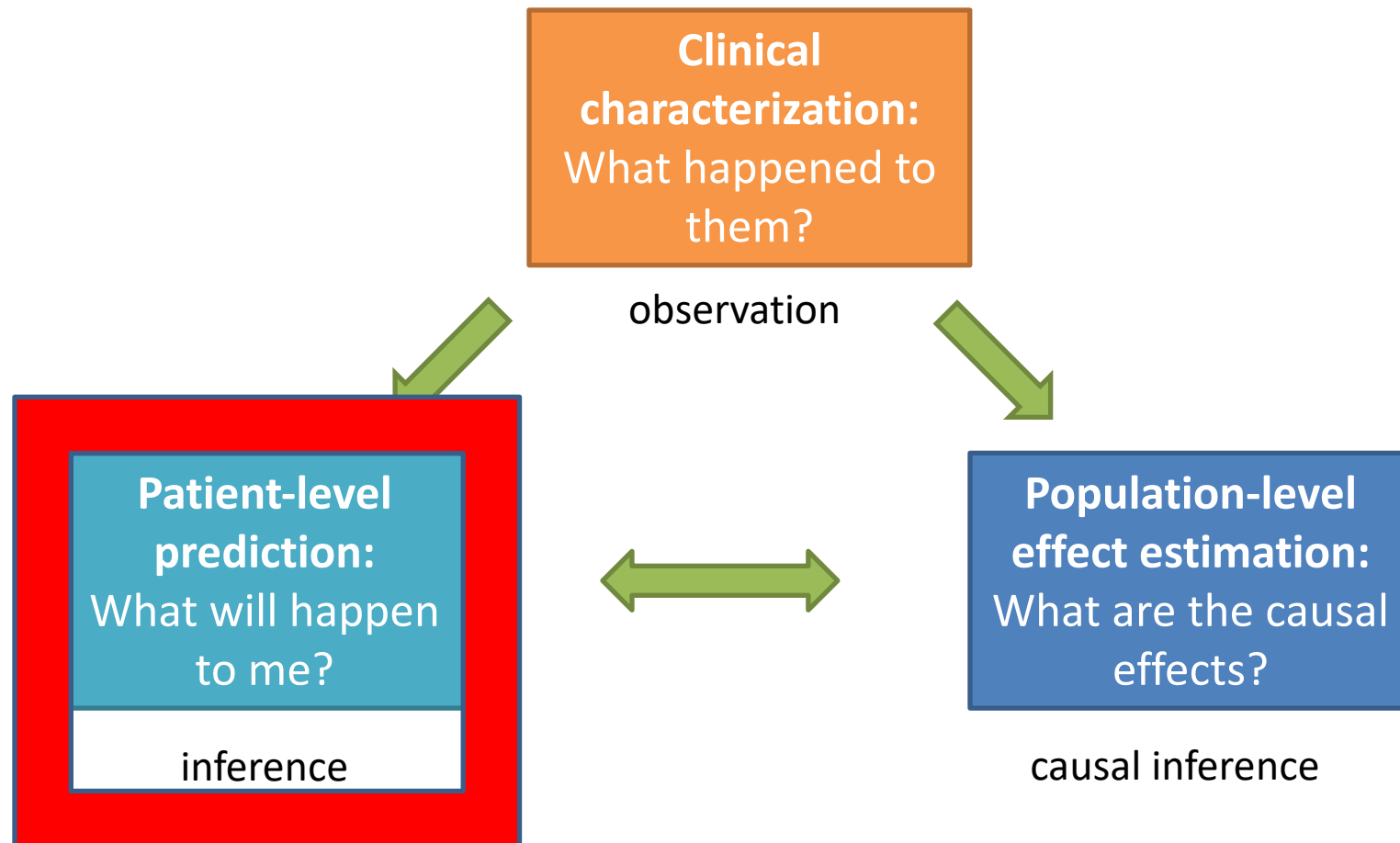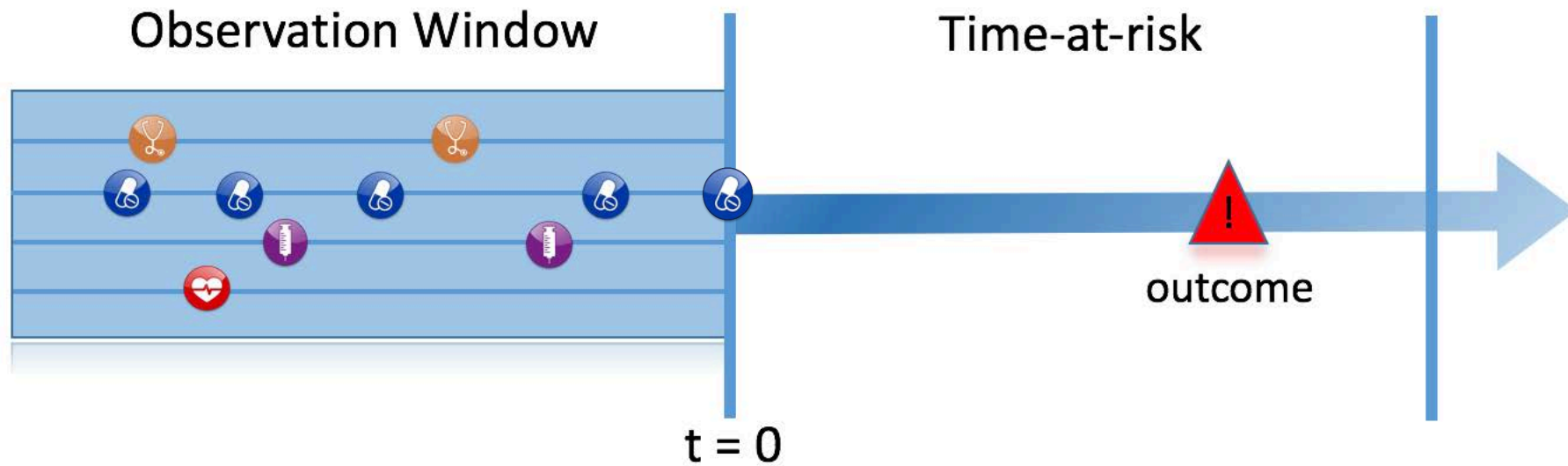
Prediction Problem Definition

Among a target population (T), we aim to predict which patients at a defined moment in time (t=0) will experience some outcome (O) during a time-at-risk Prediction is done using only information about the patients in an observation window prior to that moment in time.

Important questions to ask!

- What decision is the prediction model intended to inform?

- When is the decision made in the context of the patient's health experience and interaction with the healthcare system?

- Who is the decision-maker, and from which stakeholder vantage point are we evaluating the decision?

- What is the trade-off between True Positive, False Positive, True Negative, False Negative?

- Etc.

OHDSI Mission for Patient-Level Prediction

OHDSI aims to develop a systematic process to learn and evaluate large-scale patient-level prediction models using observational health data in a data network

Evidence Generation → Evidence Evaluation → Evidence Dissemination

OHDSI's Patient-Level Prediction Framework

JAMIA
A SCHOLARLY JOURNAL OF INFORMATICS IN HEALTH AND BIOMEDICINE

Design and implementation of a standardized framework to generate and evaluate patient-level prediction models using observational healthcare data 🔓

Jenna M Reps ✉, Martijn J Schuemie, Marc A Suchard, Patrick B Ryan, Peter R Rijnbeek

📄 PDF   ❚❚ Split View   ❝ Cite   🔑 Permissions   ❮ Share ▾

Abstract

Objective

To develop a conceptual prediction model framework containing standardized steps and describe the corresponding open-source software developed to consistently implement the framework across computational environments and observational healthcare databases to enable model sharing and reproducibility.

R-package

www.github.com/OHDSI/PatientLevelPrediction

- Vignettes
- Videos
- Online training material

Book-of-OHDSI
https://ohdsi.github.io/TheBookOfOhdsi/

Study Results
www.data.ohdsi.org

The prediction chapter and the publication are added on top of our channel in Teams

The Journey: Problem Definition

Problem Definition → Data Extraction → Training → Internal Validation → External Validation → Dissemination

**Problem pre-specification.** A study protocol should unambiguously pre-specify the planned analyses.

Transparency. Others should be able to reproduce a study in every detail using the provided information. All analysis code should be made available as open source on the OHDSI Github.

Team Effort:
- Problem Definition + Questions
- Literature Research -> Prior work, Rationale
- Study Protocol Development

The Journey: Data Extraction

Problem Definition → **Data Extraction** → Training → Internal Validation → External Validation → Dissemination
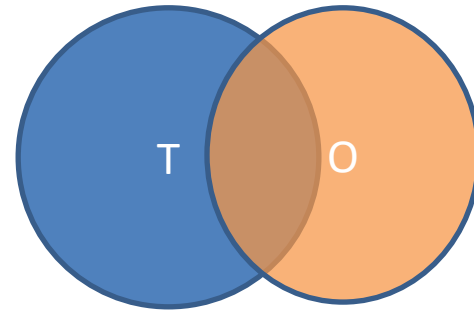
We extract data for the patients in the Target Cohort (T) and we select <u>all</u> patients that experience the outcome (O)

The Target Cohort (T) and Outcome Cohort (O) can be defined using ATLAS or custom code (see later today).

For model development all outcomes (O) of patients in the Target Cohort (T) are used.

Team Effort:
- Literature Review
- Cohort Definition

Work done with the phenotype group

The Journey: Model Development

Problem Definition → **Data Extraction** → Training → Internal Validation → External Validation → Dissemination

Data is extracted from the OMOP CDM using the Feature Extraction R-Package.

This allows for specification of the candidate predictors and time windows.
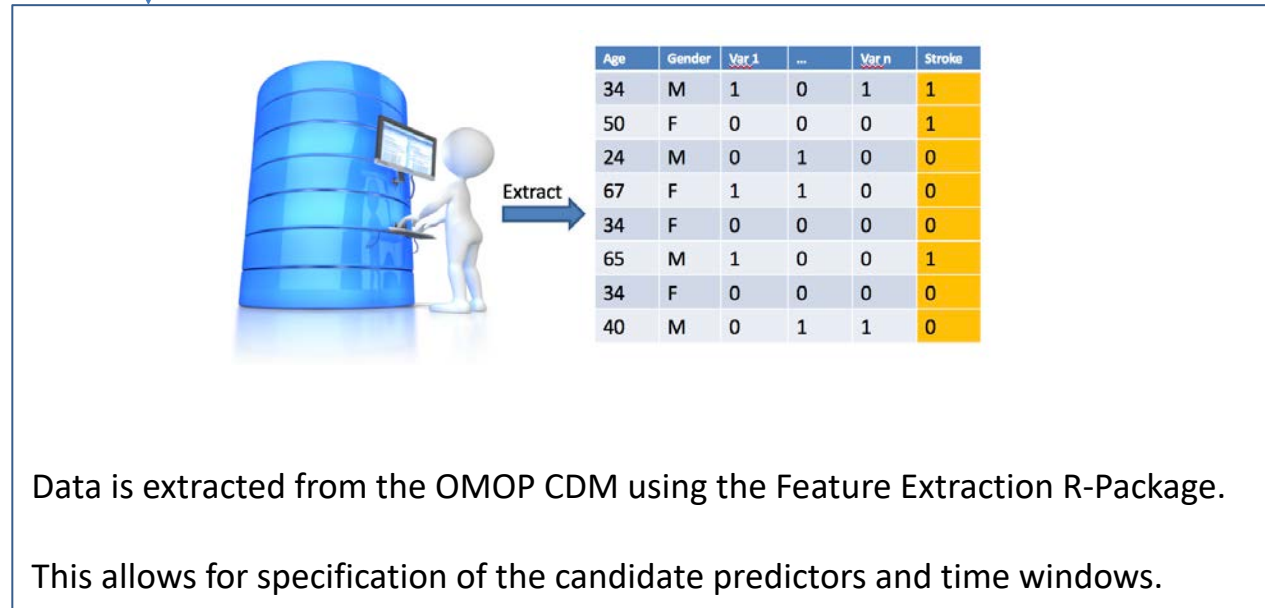
Team effort:
Cohort Diagnostics Package

Work done with other channels

The Journey: Model Development

Problem Definition → Data Extraction → **Training** → **Internal Validation** → External Validation → Dissemination

Model training and **Internal validation** is done using a train test split:

1. Person split: examples are assigned randomly to the train or test set, or

2. Time split: a split is made at a moment in time (temporal validation)

Train set | Test set

2014-01-15

- Study Package Development
- Study Execution

The Journey: External Validation

Problem Definition → Data Extraction → Training → Internal Validation → **External Validation** → Dissemination

External validation is performed using data from multiple populations not used for training.

Train

1 → Model

Apply

2
3
4

Evaluate

Auc2, Cal2
Auc3, Cal3
Auc4, Cal4
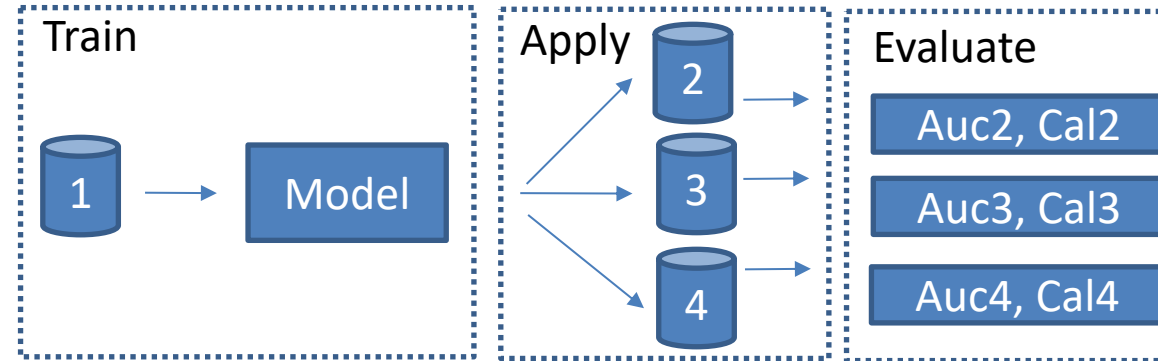
- Data Partners

The Journey: Dissemination

Problem Definition	Data Extraction	Training	Internal Validation	External Validation	Dissemination

Dissemination of study results should follow the minimum requirements as stated in the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement [1].

- Internal and external validation
- Sharing of full model details
- Sharing of all analyses code to allow full reproducibility

→ Website to share protocol, code, models and results for all databases

[1] Moons, KG et al. Ann Intern Med. 2015;162(1):W1-73

PLP Aims Study-A-Thon

Build and evaluate models developed on Flu patients to:

1) Test them on COVID patients if data becomes available

2) Have tools ready to learn on COVID patients

And,

Replicate some of the models found in literature

Team Effort

51 Participants in our channel and literature study

And Many More

Thank you all for the great collaboration in the PLP team