A framework to efficiently identify potential prognostic factors

Jenna M. Reps, PhD, M. Soledad Cepeda, MD, PhD
Janssen Research and Development, Raritan, NJ

Aim: To develop a data-driven framework that can use the OHDSI data network to combine multiple database perspectives to identify potential new prognostic factors for a specific illness

BACKGROUND

Prognostic models predict the individual risk of some future outcome during a certain time period for a cohort of people.

The primary use of prognostic models is to calculate the personalized risk of the outcome for each individual, but these models can be an excellent way to learn about the illness as they can highlight the variables that are associated to the outcome (prognostic factors).

METHOD

Data   Models   Combine   Clinical Review

1

Exploration: find important variables in each dataset using regularised logistic regression prediction models to filter useful variables.

2

Replication: check for variable importance consistency by finding all the variables that are consistent across different dataset models.

3

Clinical Review: clinician reviews variables and picks some to be further evaluated

RESULTS

Model performance for predicting 2-year risk of therapeutic knee injection and number of variables selected are presented in Table 1.

Table 1. The prediction performance on the test set.

Database	AUC	Variables with positive coefficient	Variables with negative coefficient
Truven CCAE	0.74	177	235
Truven Medicare	0.62	169	244
Truven Medicaid	0.76	105	175
Optum	0.75	166	258

There were 41 variables associated with an increased risk and 55 variables associated with a decreased risk selected consistently across datasets (Table 2).
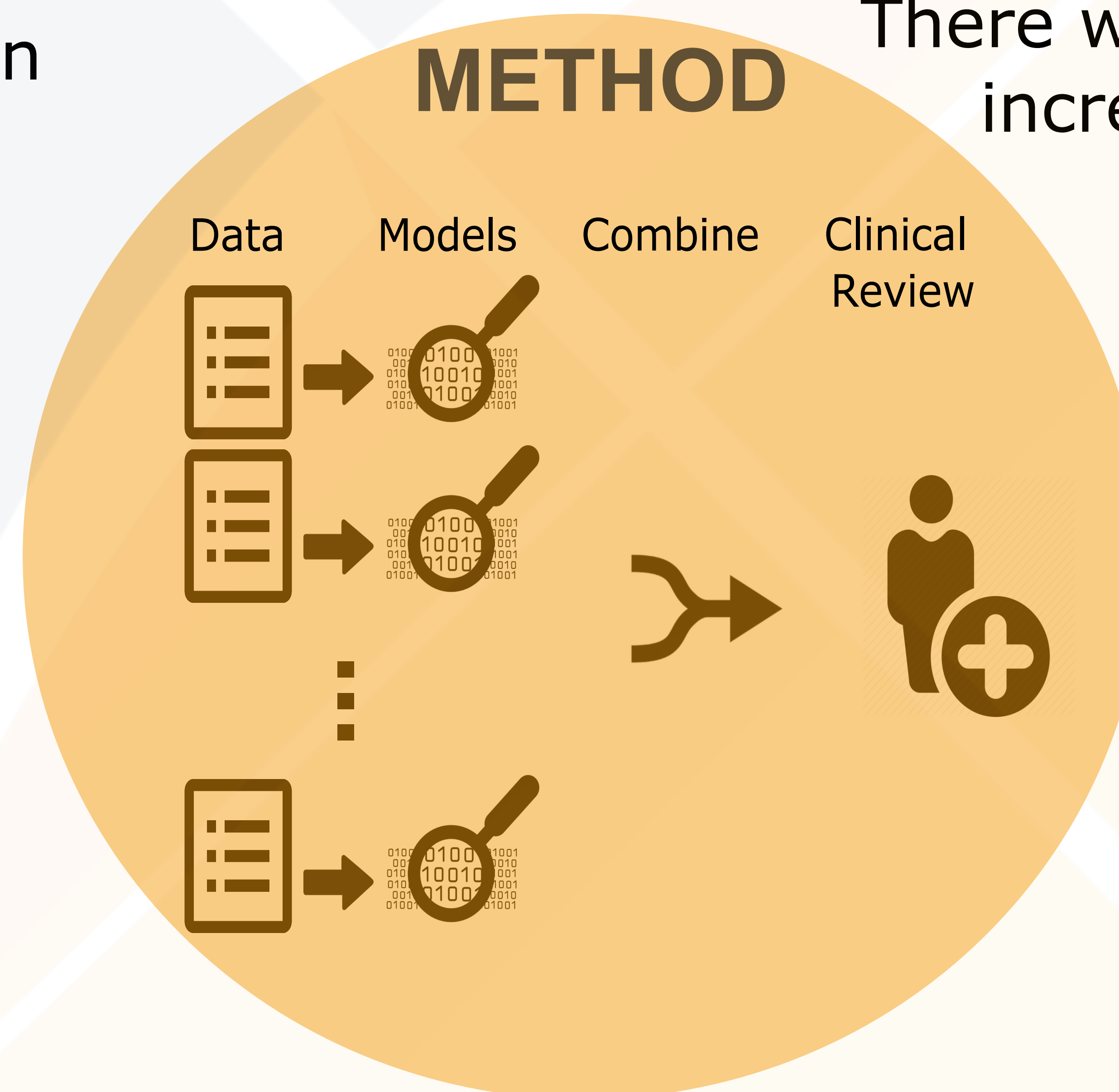
Table 2. The prognostic factors identified by applying the framework, known prognostic factors are in bold.

Variable	Direction
Appetite and general nutritional disorders (Obesity)	**Increased risk**
Malabsorption conditions	Increased risk
Procedure -Cardiac and vascular investigations (excl enzyme tests)	Increased risk
Venous varices	Increased risk
Mood disorders and disturbances NEC	Increased risk
Ingredient group with parent: ANTIGOUT PREPARATIONS	**Increased risk**
Iron and trace metal metabolism disorders	Increased risk
Gastrointestinal infections	Increased risk
Bone disorders (excl congenital and fractures)	Decreased risk
Autoimmune disorders	Decreased risk
Musculoskeletal and soft tissue investigations (excl enzyme tests)	Decreased risk
Psychiatric disorders NEC	Decreased risk
Bone, calcium, magnesium and phosphorus metabolism disorders	**Decreased risk**
Gender = MALE	**Decreased risk**
Age group: 35-39	Decreased risk
Thyroid gland disorders	Decreased risk
Lipid metabolism disorders	Decreased risk
Aneurysms and artery dissections	Decreased risk

CONCLUSION

These new prognostic factors can be reviewed by a clinical expert to gain insight into the mechanism behind the illness progression.

In future work this study should be replicated across the OHDSI network and the identified potential risk factors could be evaluated using estimation methods to determine whether they are causal or not.

Want to be involved in continuation of this study?

Janssen
PHARMACEUTICAL COMPANIES
OF Johnson&Johnson