

# HERO: Human Extraordinary Robust Organism

Mina Bjelogrić, Julien Ehram, Christophe Gaudet-Blavignac, Christian Lovis

## SEMI-AUTOMATIC MATCHED COHORTS SELECTION

The dataset include several cohorts categorized by their inclusion criteria:

- based on **biological alteration**
- based on **diagnostic** criteria;
- based on **environmental or behavioral risk factors**.

Cases will be selected when patients do not suffer from an expected consequence.

### Example of biological alteration: Dyslipideamia

Example of expected consequences :

- risk factor for atherosclerosis
- coronary artery disease



### Example of diagnostic: Diabetes mellitus

Example of expected consequences:

Cardiovascular risk	Prevalence
Coronary heart disease	14-21%
Heart failure	19-26%
Peripheral artery disease	16-29%
Stroke	9-12%
Retinopathy	34%
Neuropathy	31-74%*
Nephropathy	29-61%

### Example of environmental or behavioral risk factor : Smoking

The expected consequences for patients exposed to smoke and the major consequences can be summarized as:

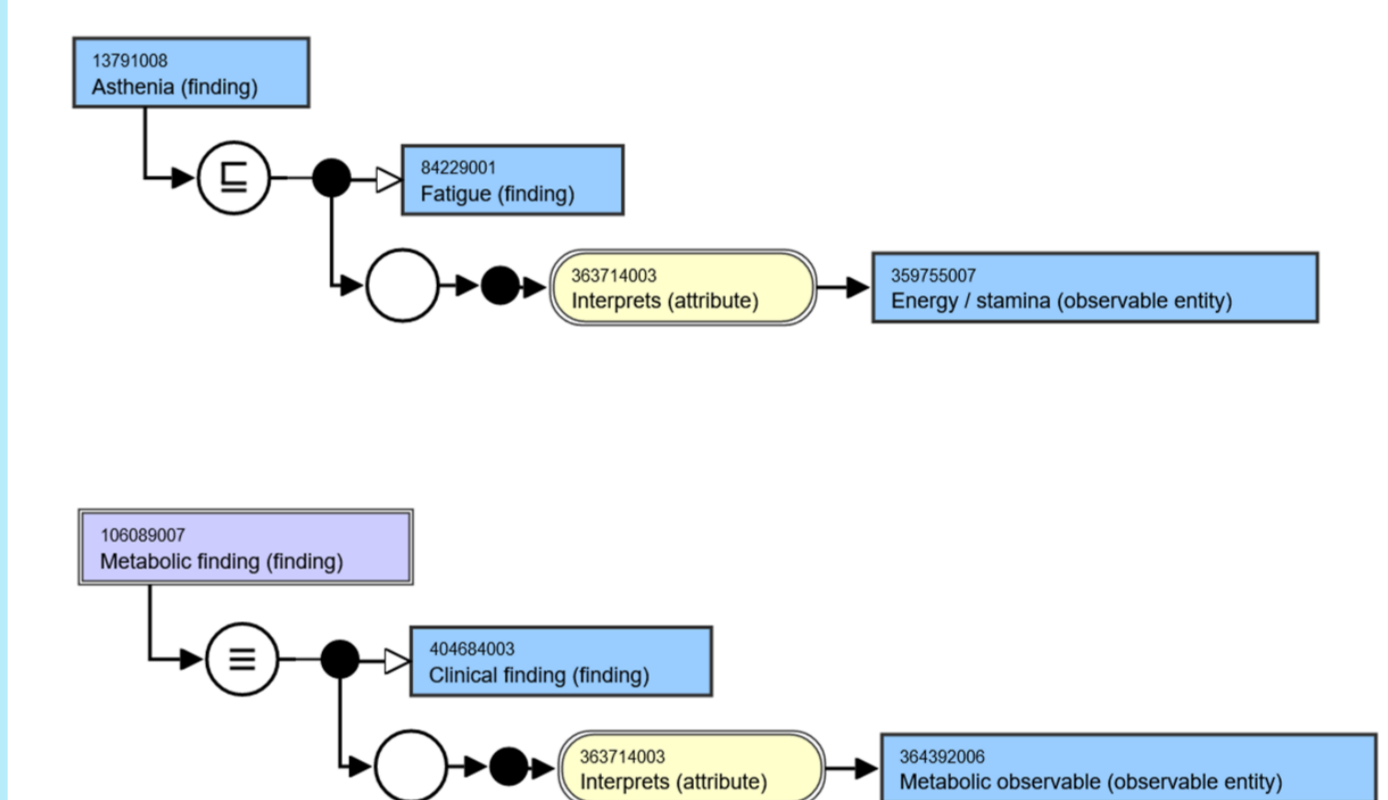
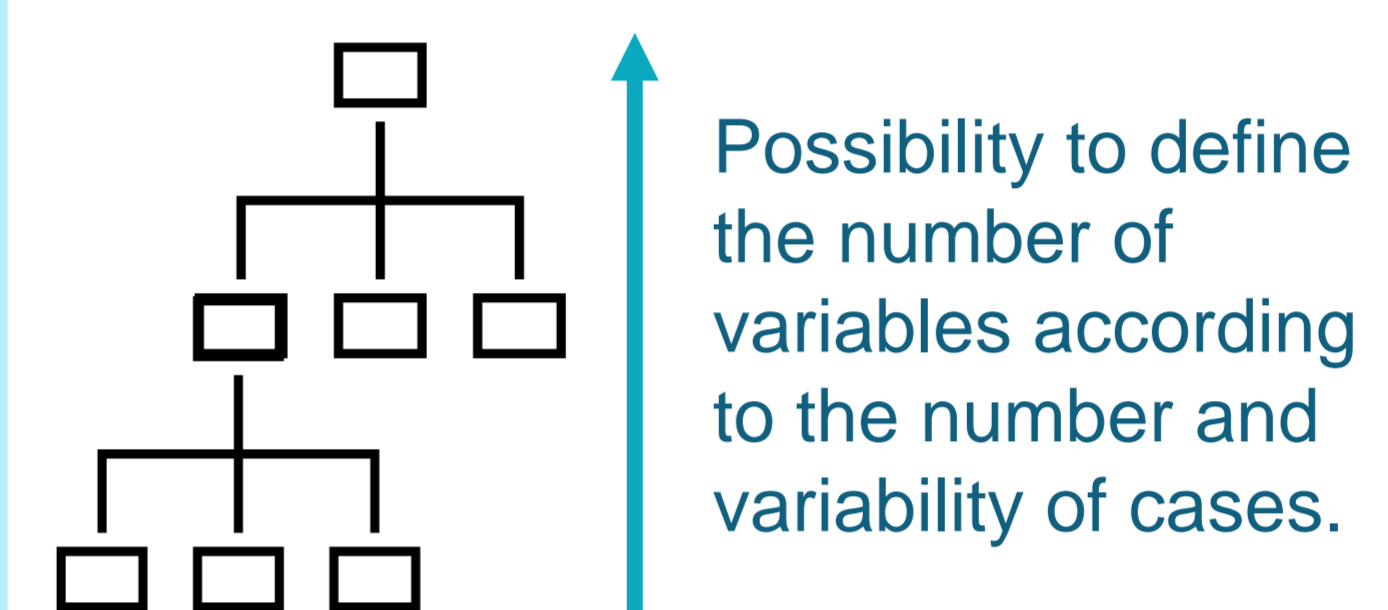
- Lung cancer
- Upper aerodigestive cancer
- Other cancer
- Other respiratory diseases
- Cardiovascular diseases

\* In people with diabetes mellitus, type II

## UNKNOWN PROTECTION MECHANISMS DISCOVERY

### Knowledge based dimensionality reduction

This is based on work done by Prof.Lovis' research group and consists in adding semantically enriched metadata to the variables of the HUG clinical database (DPI). Experts in the various fields, such as physicians, nurses, biologists, take the variables used in the DPI, such as a "sensation de faiblesse", and encode it using SNMOED-CT (SCT).



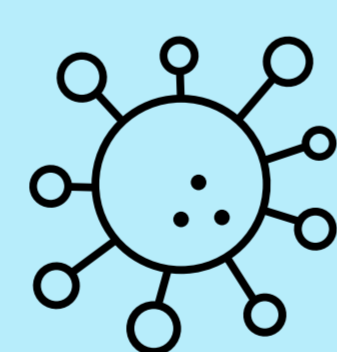
Medicine is historically built on pathological processes and their corrections.



The proposed methodology focus on a different and innovative approach centered on the analysis of protective factors.



Protective factors are associated with a lower likelihood of a negative outcome and linked with positive features that we could call "shields".



### Patient representations and rule induction models

- Low dimensional medical concept representations
- Massive correlation matrix for different concepts levels
- Rule induction methods for interpretable discoveries based on data-driven representations of cases.

### Research impact

This project goals are:

- a novel data-driven methodology for the semi-automatic selection of patients who have a higher probability of exhibiting unknown protection mechanisms and their matched cohorts → increase probability of finding novel protection mechanisms, versus discoveries historically made "by chance".
- proposing interpretable models by design with neuro-symbolic rule learning systems. → propose protection scores.

