

# Introduction

A previous proof of concept of economic burden study on French hospital database confirmed how data mining algorithm helped identifying specific HIV patient profiles over-expressing costs during hospitalization.

## Objectives

The objective of the present study is to find the **profiles of patients** living with HIV (PLHIV) who overexpress the **cost** of care, using all reimbursement data of the French health insurance (SNIIRAM database) covering both hospital and community care management.

## Methods

**Design: A retrospective cohort study on the SNIIRAM**

SNIRAM contains individualized data on all reimbursed health expenses for the French population. Annual cost per patient was assessed for the economic burden.

## Machine learning algorithm

**A Boosted Decision Tree regressor** was used to find profiles overexpressing costs (a profile is a combination of several variables). Algorithm was customized to match the size and complexity of health data.

## Input variables

Gender, Age, Geographical place of care,  
Type of social fund, Disadvantaged index

**Selection: validated algorithm or medical review**

- Cardio-neurovascular diseases
- Antihypertensive treatment
- Lipid-lowering treatment
- Diabetes
- Cancer
- Psychotropic medications
- Degenerative disorders
- Neurological disorders
- Chronic respiratory disease
- Chronic inflammatory disease
- Chronic end-stage kidney failure
- Liver or pancreas disease

## Conclusion

- In-depth uses of large healthcare databases require innovative analysis methods.
- Combining an epidemiological study methodology (cohort definition and monitoring, medical review, economic burden) with an innovative mathematical model (Machine Learning) to obtain accurate results regarding patient profiles.
- Profiles are usable to objectify action plans to improve care policies.

## Analysis methodology

