

A new analysis methodology to find atypical cost profiles in patients living with HIV from a national claim insurance data (SNIIRAM)

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

SNIIRAM 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



Raw Data

Cohort extraction from SNIIRAM

98,000 patients

Inclusion and Follow-up

Period: 1st Jan. 2013 ➤ 31 December 2013
Criteria: presence of a care related to HIV
Follow-up: 1 year after the inclusion

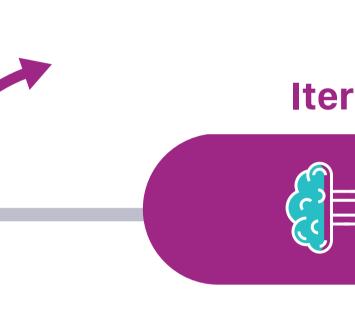


Processed data + Target

Building input variables

Geographical and social Insurance scheme HIV comorbidities

Target = Cost of hospital and community care from National Health Insurance



Iterative refinement of parameters

Machine Learning

Building Machine Learning algorithm

A supervised learner that automatically select the most relevant variables for a meaningful classification of individuals. The objective of the algorithm is to identify patient profile with atypical costs.



Results



Profile 1

31,950 (50.1%)

Low costs [0 - 10€K]

Key caracteristics

Age between 15 and 45 years old

Not treated with Lipid-Lowering drugs



Profile 2

7,900 (12.4% patients

High costs [15 - 80€K]

Key caracteristics

Age above 45 years old
Treated with Lipid-Lowering

drugs



Profile 3

36,850 (37.6%) patients

Medium cost [10 - 15€K]

Key caracteristics

Age above 35 years old

No psychotropic treatment

No psychiatric disease
No liver, nor pancreas
disease



