INTRODUCTION

Data mining is not new and has been efficiently applied in various domains (Bank, Cybernetics, Marketing, Energy, etc.). The amount of data collected in medical information systems is tremendous, but much data remains unused because of their complexity. This huge increasing volume of data requires new analytical approaches that are efficient, sensitive and better than classical statistical to handle Big Data.

OBJECTIVE

The main objective of this study is to assess the capability of a state-of-the-art data mining technique to be applied on a healthcare database in order to better understand drivers of healthcare expenditure and the management of diseases.

METHODS

KEY FIGURES

- Number of patients: 30,294
- Number of days: 70,180
- Total annual cost: €632 per patient

DATA EXTRACTION

The PMI-4502 (French Medical Information System – Medicine, Surgery, Obstetric unit) database was used to extract all hospital stays in 2013 with at least one of the following HIV ICD-10 (International Classification of Diseases, 10th revision) codes as principal diagnosis: B20*, B21*, B22*.

RESULTS

- 1-year follow-up of patients & classifications of stays
- Explanatory approach, discover the unknown with no preconception
- Logistic regression tells you that X% of the cost is explained by having this or this

ADVANCED DATA MINING APPROACH

Prediction of costs associated with the hospital management of HIV patients in France

CONCLUSION

The current study is a pilot that successfully demonstrates how a data mining technique, from the field of Artificial Intelligence, can help better understand hospital costs for HIV patients: our data mining algorithm identified specific patient profiles which explain the differentiating cost drivers in HIV-inpatient care. Classical statistical approaches would struggle to provide such detailed profiles with numerous combinations of variables.