Elsa Duteil[1], Romain Lorenzi[2], Astrid Pagniez[1], Ludovic Larmarsalle[2]

[1] Novartis Pharma S.A.S., Rueil-Malmaison Cedex, Paris, France, [2] HEVA, Lyon, FRANCE

Study of the severity of vaso-occlusive crisis for patient suffering from sickle cell disease in France
using a Machine Learning approach on the French medical information system database

Introduction

A previous study described the population of patients hospitalized for **sickle-cell disease** (SCD) in France. The proportion of patients experiencing Vaso-Occlusive Crisis (VOC), hospital events related to SCD, was estimated. Here, we show an **analytic approach** to detect and **predict** patient profiles the more at risk of experiencing it.

Objectives

◆ Assessing if a **Machine Learning algorithm** can discover hidden values in health data

◆ Bringing a better understanding of SCD: **predict** the severity of Vaso-Occlusive Crisis

Methods

Study design

◆ A **retrospective** cohort study on the 2014 national hospital discharge database – all SCD patients.

◆ **Follow-up**: 1 year before and 2 years after inclusion (2013 - 2016)

Target

The **presence** of VOC is expressed by:

◆ The **frequency** of VOC (annual = number of VOC)

◆ The **average time between** two VOC.

Machine learning algorithm

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

Conclusion

◆ In-depth uses of large healthcare databases require **innovative** analysis methods.

◆ Combining an **epidemiological study** with an innovative **mathematical model** (Machine Learning) brings accurate results regarding patient profiles.

◆ More medical data regarding patients' health status would allow **high quality predictive models.**

Results

5,630 patients

7 general variables (age, gender, …)

1 variable related to the pregnancy status

4 variables related to the 2-years follow-up

35 input variable

3 variables related to patients' medical history

20 comorbidities

2 — 2 targets for the presence of VOC

Top 5

1,808 — General symptoms and signs

1,083 — Influenza and pneumonia

909 — Anemia and aplasia outside

689 — Symptoms and signs related to circulatory and respiratory systems

490 — Pain in limb

Building a Machine Learning algorithm

A supervised learner implemented in **Python**, from **Scikit-learn** scikit-learn.org

TOP 2 PROFILES among 20

364 patients 6,5%	254 patients 4,5%
HIGH VOC Frequence (> 4)	**LOW VOC Frequence (< 0,5)**
KEY characteristics	**KEY characteristics**
Transfusions the 1st year: **< 6.5**	VOC before 2014: **< 14**
Transfusions the 2nd year: **< 1.5**	Transfusions the 2nd year: **< 0.5**
	LOS when hospital stay for VOC before 2014: **< 1 day**
	Total time at hospital (2013-2014): **< 1 day**

NOVARTIS

HEVA