



Histoire d'un appariement avec un SNDS Fils

Meet up HDH

28 Novembre 2019

Un soir d'automne 2017... Martine presque Présidente !



Pr Martine **Gillard**

Chef de service de cardiologie - CHU Brest
Présidente Société Française de Cardiologie

« Je suis présidente dans 3 mois. J'ai deux ans de présidence et deux objectifs:

1. L'Europe
2. Les registres et la recherche

La science et la formation sont les missions de notre société savante.

Il faut que nous arrivions à pérenniser les registres, et améliorer la qualité des données tout en réduisant le temps passé par les cardiologues. Trouvez une **solution**, France TAVI sera le pilote. »

Le projet pilote: France TAVI

Promue par la SFC à la demande de la HAS

2010

2012 - 2013

2018

France 2
4.201 patients
100% MONITORÉ

France TAVI
28k patients

France TAVI
RHU StopAS
12k patients

30 publications



Volume 68, Issue 15, October 2016
DOI: 10.1016/j.jacc.2016.07.747

Env. 45'000

patients porteurs d'un TAVI

Env. 50

Centres implanteurs

Env. 10'000

Nouveaux patients par an

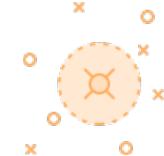
Le modèle d'hier...



Données hospitalières

- Antécédents
- Facteurs de risques
- Etat cliniques
- Données de procédures
- Complications hospitalières

eCRF



Suivis annuels

- Statut vital
- Hospitalisation & évènements MACCE
- Consommation de soins
- Traitements médicaux
- Données échographiques

eCRF

...que nous voulons modifier...



Données hospitalières

- Antécédents
- Facteurs de risques
- Etat cliniques
- Données de procédures

Système Information Hospitalier

Complications hospitalières



Suivis annuels

- Statut vital
- Hospitalisation & évènements MACCE
- Consommation de soins
- Traitements médicaux



Données échographiques

eCRF

On rencontre la CNAM, pour la méthode d'appariement

Direct

Je dispose du NIR

J'apparie via le NIR « FOINisé » avec le SNDS

Déterministe

Je dispose de variables qui permettent de « reconstruire » le NIR

Je passe par la CNAV, un premier appariement permet de récupérer les NIR via le Système National de Gestion des Identités.

J'apparie via le NIR « FOINisé » avec le SNDS

Probabiliste

Je ne dispose ni du NIR ni des variables qui permettent de le « reconstruire »

Quelles sont les variables présentes dans le SNDS qui me permettent de rapprocher un patient [Cohort] du même patient dans le SNDS ?

Bonne nouvelle, nous avons un acte traçant: DBLF001 ou DBLA004

Target: 90%

On architecture une première plate-forme de traitements

Bulle CNAM

Tropisme CRO

Statistiques « classique »

SAS

Pas de compétences
infrastructure et sécurité

HADS & SNDS Fils

Tropisme Tech

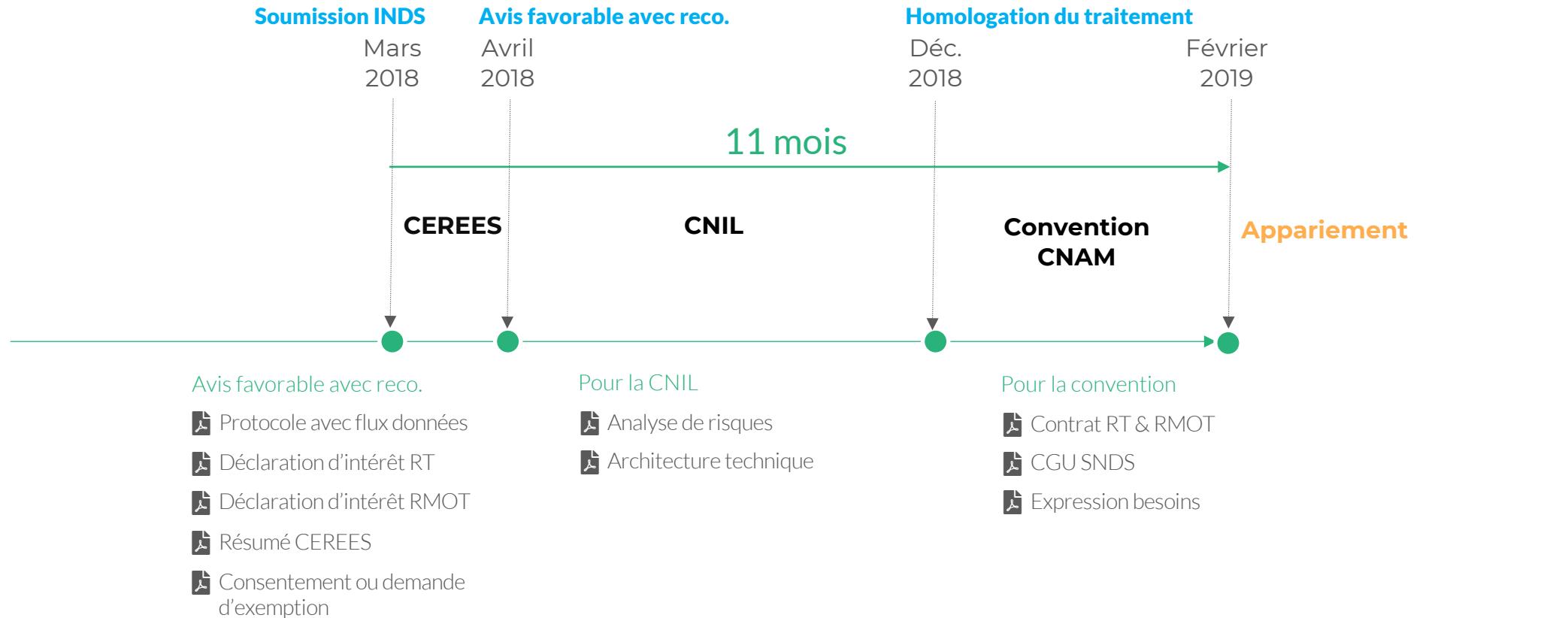
Modèles prédictifs
Machine Learning
Process mining

Python
Scala
Spark

Compétences
infrastructure et sécurité

Et Go... On se lance dans la procédure d'accès aux données

Découverte pas à pas...



Un schéma d'infrastructure

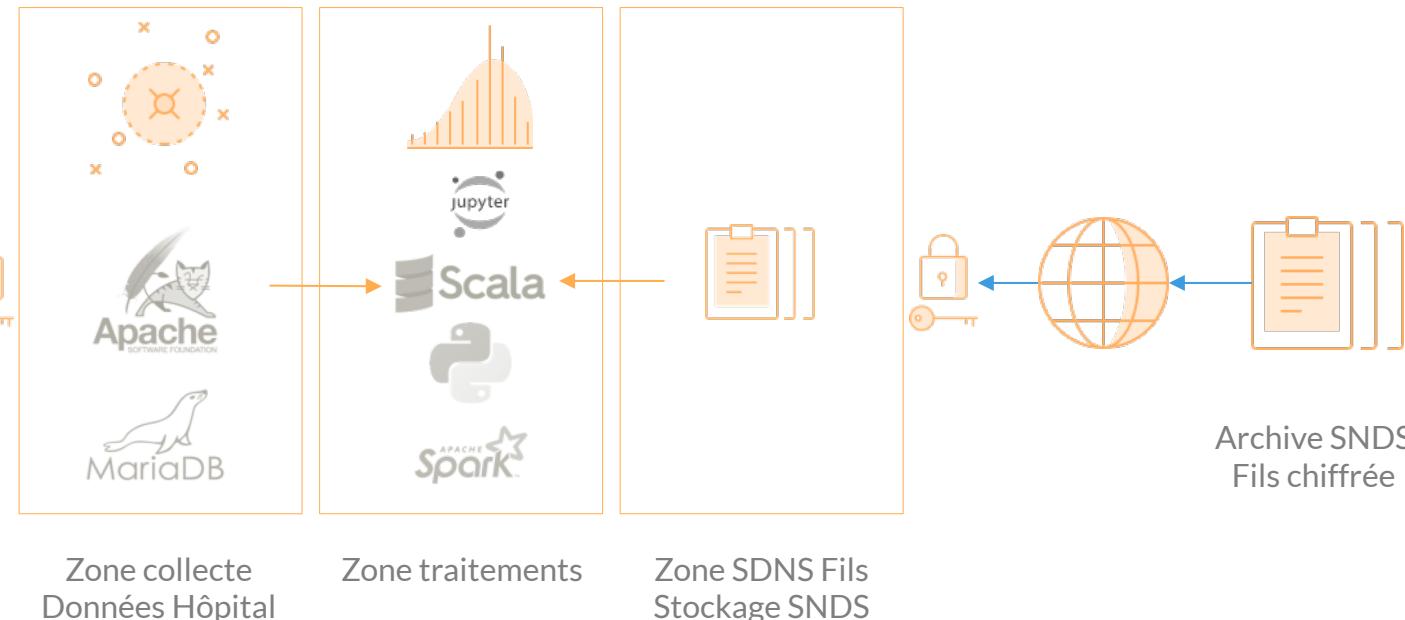
Une analyse de risques eBIOS, un HADS et un peu de travail

Hôpital



Clinityx

Au total 8 serveurs chez un HADS



CNAM

Etape 1: L'appariement probabiliste

Pas si simple...

Appariement test

97.6%

Appariement principal :

- BEN_SEX_COD (Sexe)
- BEN_NAI_ANN (Année de naissance)
- BEN_NAI_MOI (Mois de naissance)
- ETA_NUM (Établissement)
- EXE_SOI_DTF à +/- 3 jours (Date de soin)

&

Appariement secondaire :

- TOP EXE_SOI_DTD à +/- 3 jours
- TOP EXE_SOI_DTF égaux
- TOP BDI_COD égaux
- TOP Code CCAM égaux

Appariement 1

84%

Appariement 2

91.6%

Appariement 3

83.4%

Appariement final

Appariement principal et strict:

- o Acte CCAM du séjour (DBLF001 ou DBLA004)
- o Sexe du patient
- o Année de naissance du patient
- o N° FINESS de l'établissement
- o 3 jours de différence maximum entre la date de sortie de l'hôpital du fichier Excel et celle présente dans le PMSI

&

Appariement secondaire (sous forme de TOP) :

- 3 jours de différence maximum entre la date d'entrée à l'hôpital du fichier Excel et celle présente dans le PMSI
- Dates de sortie de l'hôpital égales
- Mois de naissance égaux

a presque atteint l'objectif

89.9%

De la population totale est appariée
Implantation du 01/02/2010 au 31/12/2017

30'913

patients appariés

12 ans de données de 30'913 patients dans quelques CSVs

Renouvellement du SNDS tous les ans en Juillet

1049

Fichiers CSV



Données SNDS
Date d'implantation

-2 ans



Données eCRF
France 2 ou France TAVI



Données SNDS
Date d'implantation

+10 ans

Etape 2: Une longue phase de **Data engineering**

Toujours un peu en cours...

Six mains et quatre outils:

- Dictionnaire de la DRESS
- Confluence
- Python
- Scala

Pour obtenir 3 bases de données Parquet:

- Statut vital des patients
- Hospitalisation des patients
- Traitements médicamenteux des patients

Et tenter de comprendre comment tirer le meilleur de la fusion des données cliniques et des données du SNDS.

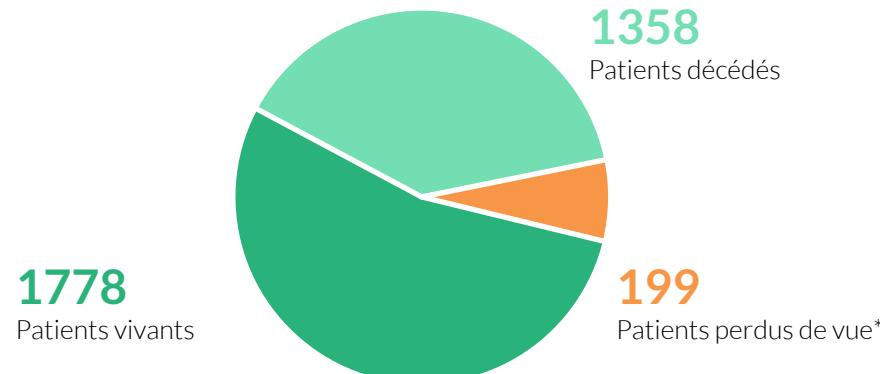
Objectifs:

- Comprendre comment utiliser de manière optimale les données du SNDS
- Apprendre à fusionner les données cliniques et médico-administratives

Pour une étude monitorée, le SNDS enrichit un peu...

Population de France 2

Source de données
eCRF



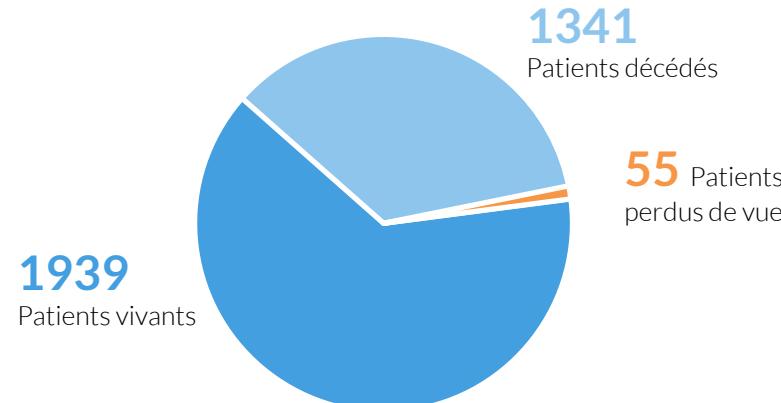
Patients perdus de vue= Patients sans donnée lors du suivi en cours

Suivi à 3 ans

3'335

patients implantés
en 2010 - 2012

Source de données
SNDS

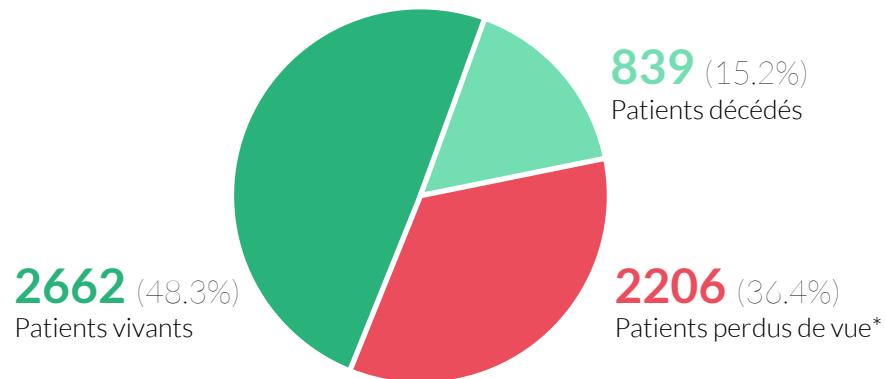


Patients perdus de vue= Patients sans trace de consommation de soin dans le SNDS dans l'année qui précède

Pour des registres en revanche, no comment !

Martine est très contente !

Source de données
eCRF



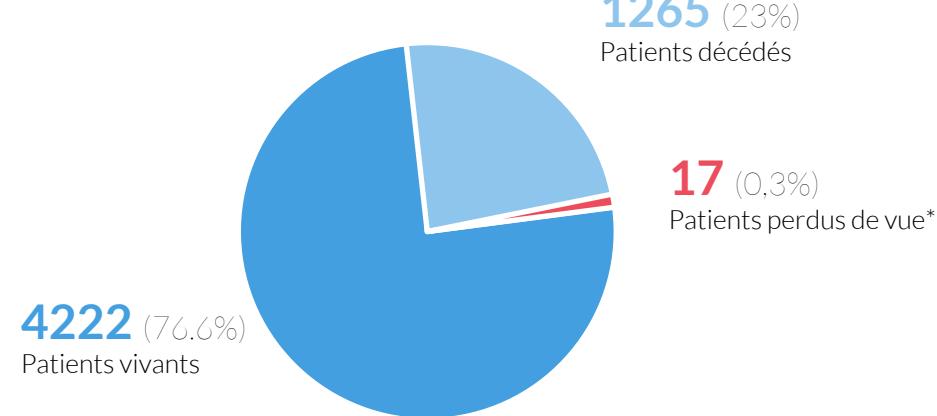
Patients perdus de vue= Patients sans donnée lors du suivi en cours

Suivi à 2 ans

5'507

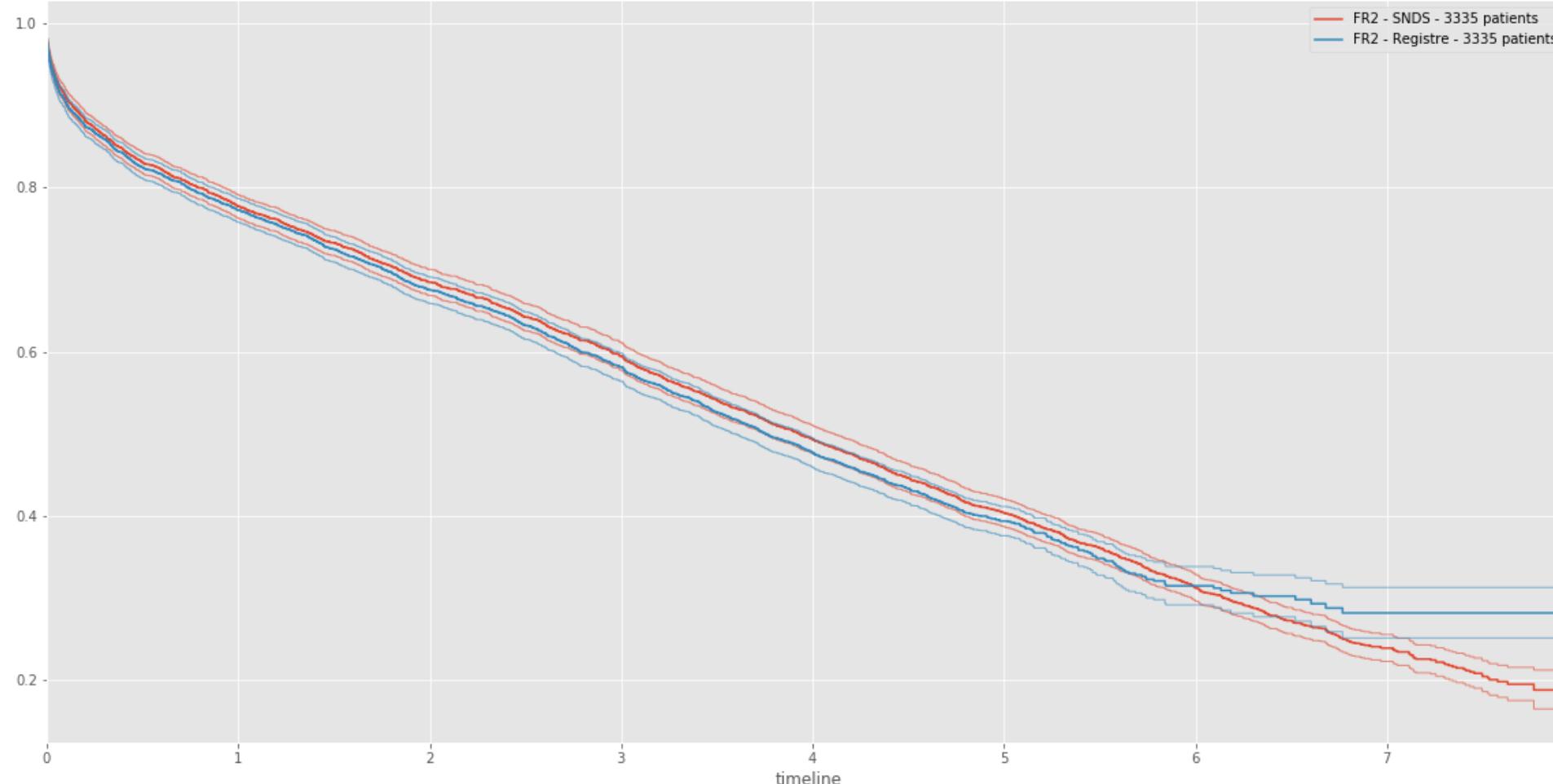
patients implantés
en 2015

Source de données
SNDS



Patients perdus de vue= Patients sans trace de consommation de soin dans le SNDS dans l'année qui précède

Comparer et allonger les courbes de mortalité



Les morts c'est fait, mais pour les évènements indésirables...

Exemple de l'insuffisance cardiaque

On cherche à comparer le nombre d'hospitalisations pour IC dans l'année qui suit l'implantation du TAVI:

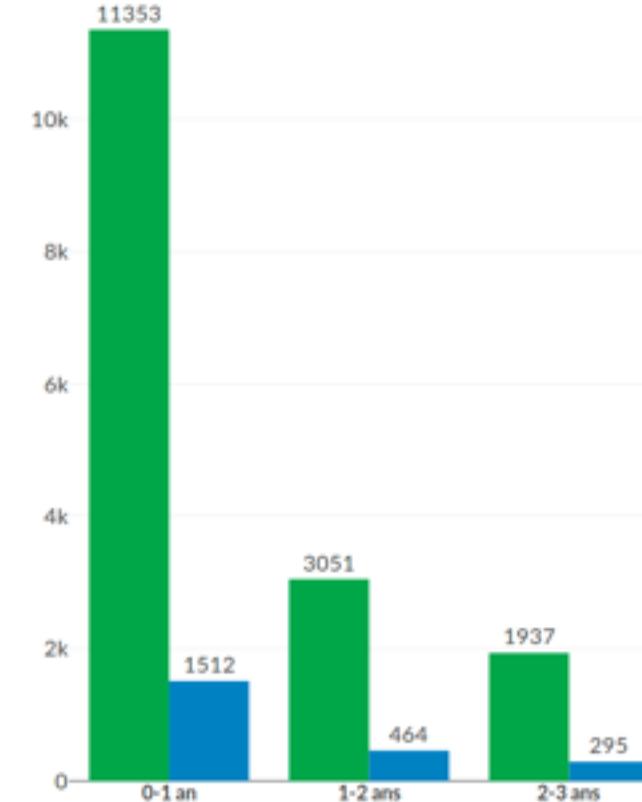
- France TAVI: 1512
- SNDS: 11'353

Pour la cardiologue, l'IC est grandement surévaluée dans le PMSI en raison de son attractivité économique

Pour la CNAM, la cartographie des pathologies est très large

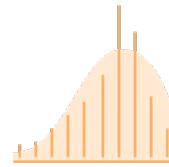
Notre plan pour connaître la vérité, on part de France 2 qui a été monitorée:

- On identifie tous les évènements indésirables conduisant à une hospitalisation dans France 2
- On identifie pour chaque évènement France 2, l'hospitalisation correspondante dans le SNDS
- On identifie au sein de chaque hospitalisation les CIM10, LPP et CCAM associés
- On détermine si l'hospitalisation SNDS non déclarée dans France 2 correspond à l'évènement indésirable au sens de France 2
- On évalue l'effet centre et année de codage
- On développe un algorithme par type d'évènement



Bref on commence à exploiter les données...

En cours



Actualisation de la publication
New England de France 2

En cours



Identification des **facteurs prédictifs de mortalité** précoce

4442 variables en paramètres

Comparaison Cox et Méthodes de ML

Besoin d'établir des classifications pour des données plus qualitatives

