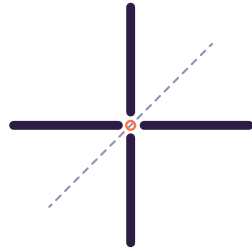




Big Data en santé du discours aux applications pratiques

Les acteurs du Big data en santé
Alexandre Vainchtock

Potentiel d'utilisation des données du SNDS



Médico-économie

Réalisation d'études portant sur l'évaluation du fardeau économique d'une pathologie, d'une prise en charge thérapeutique, sur la consommation de soins et leurs coûts (ville, hôpital, indemnités journalières)



Epidémiologie et pharmacovigilance

Réalisation d'études portant sur l'épidémiologie d'une pathologie ou les événements indésirables d'un produit de santé.

Potentiel d'utilisation des données du SNDS



Parcours et offre de soins

Parcours de soins :

Réalisation d'études portant sur l'analyse des parcours thérapeutiques des patients, centrés sur le recours aux différents acteurs de santé sur une période donnée.

Offre de soins :

Réalisation d'études portant sur l'analyse de l'offre de soins, centrées sur les professionnels de santé, les établissements de santé, les services d'urgence, les laboratoires de biologie médicale et les officines



Innovations méthodologiques

Réalisation d'études portant sur de nouvelles méthodologies appliquées au champ de la santé.

Application de nouvelles disciplines analytiques avec le machine learning, le process mining.

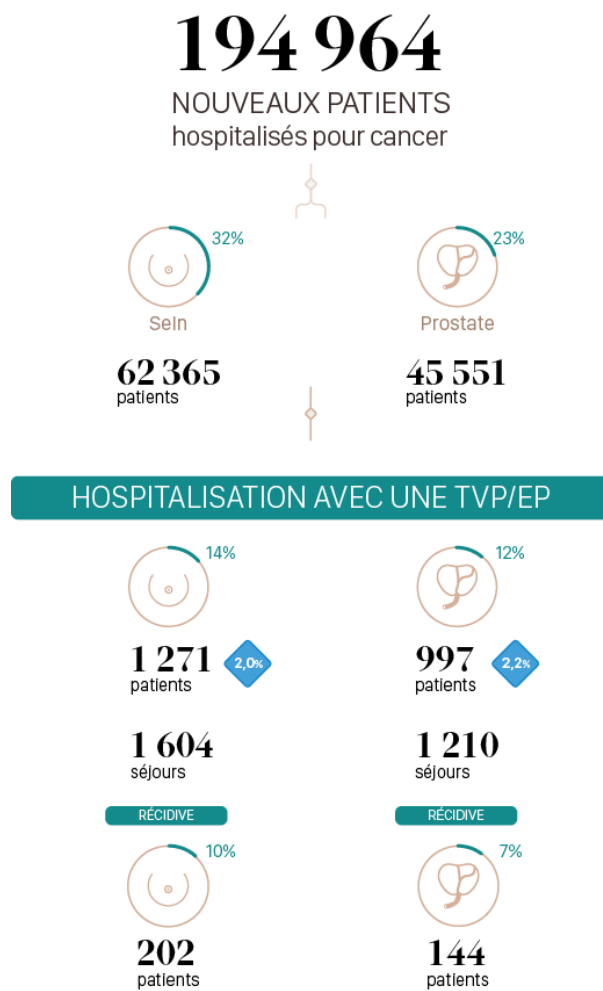
1

Applications au sein de la médico-économie

Du burden of disease à la modélisation d'impact budgétaire

Analyse des TVP-EP

patients atteints de cancer du sein et de la prostate



Analyse des événements TVP-EP auprès des patients hospitalisés pour cancer du sein et de la prostate



POPULATION

Patients traités pour un cancer du sein et de la prostate en 2010



FOLLOW-UP

Suivi prospectif de 2 années calendaires



DATA

Nombre d'hospitalisations pour TVP-EP

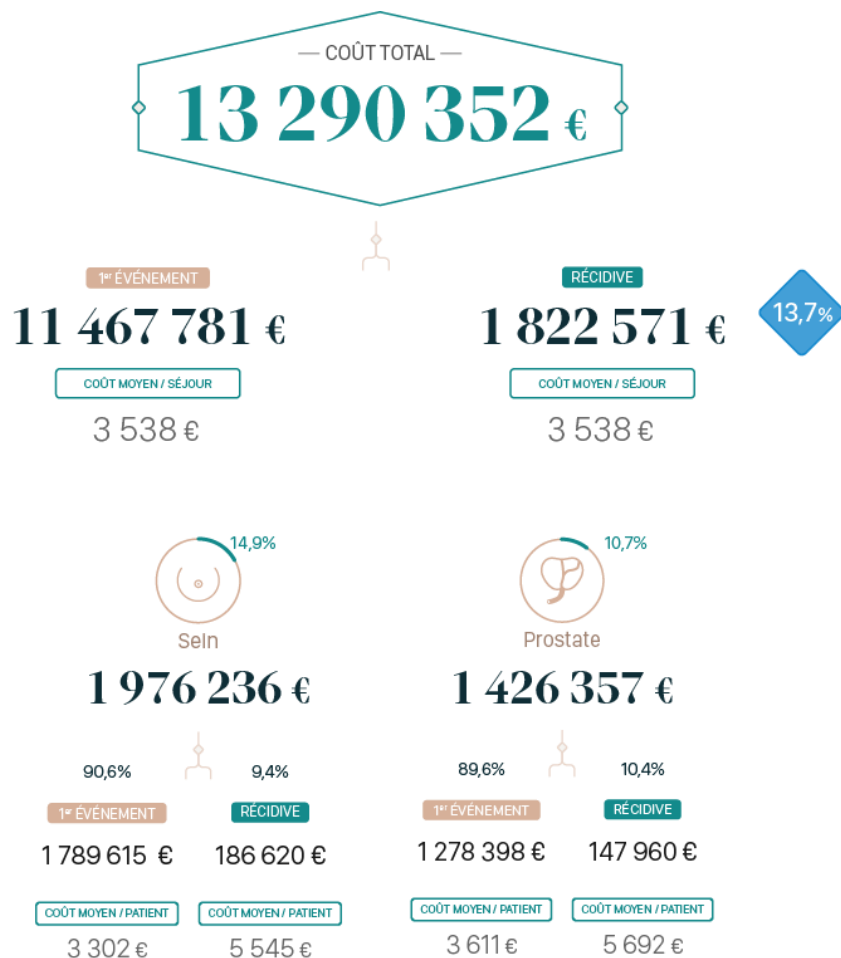
NOMBRE DE RÉCIDIVES SUR 2 ANS

202 CANCER DU SEIN

144 CANCER DE LA PROSTATE

Coût des TVP-EP

patients atteints de cancer du sein et de la prostate



Analyse des événements TVP-EP auprès des patients hospitalisés pour cancer du sein et de la prostate



POPULATION

Patients traités pour un cancer du sein et de la prostate en 2010



FOLLOW-UP

Suivi prospectif de 2 années calendaires



DATA

Nombre d'hospitalisations pour TVP-EP

COÛT TOTAL DES TVP-EP SUR 2 ANS

1,98 M€

CANCER DU SEIN

1,43 M€

CANCER DE LA PROSTATE

2

De l'épidémiologie à la pharmacovigilance

Auprès d'une population en conditions réelles de prise en charge

Cohorte de patients atteints
d'un KC du poumon



HEVA



COHORTE ANALYSEE

FOLLOW-UP

INDICATEURS SOCIO-DÉMOGRAPHIQUES

RÉSULTATS

jeudi 29 mars 2018

Risques ischémiques et hémorragiques liés aux gammes d'endoprothèses (stents) coronaires en France

Etude à partir des données du SNIIRAM

**Rapport d'étude Janvier 2018 HAS/ ANSM*

OBJECTIF :

Quantifier en « vie réelle » les risques ischémiques et hémorragiques et les différences entre les stents actifs et les stents nus/non actifs



MÉTHODE :

Sujets ayant eu une implantation de stent en 2014 âgés de plus de 18 ans et affiliés au régime général



Le risque de survenue à 1 an a été comparé entre gammes en utilisant un modèle de Cox ajusté sur les caractéristiques à l'inclusion

SYNTHESE :

Pas de différence de caractéristiques sociodémographiques, médicales entre les gammes



Après ajustement sur les caractéristiques à l'inclusion et prise en compte de l'effet centre, les risques n'étaient pas différents entre les gammes



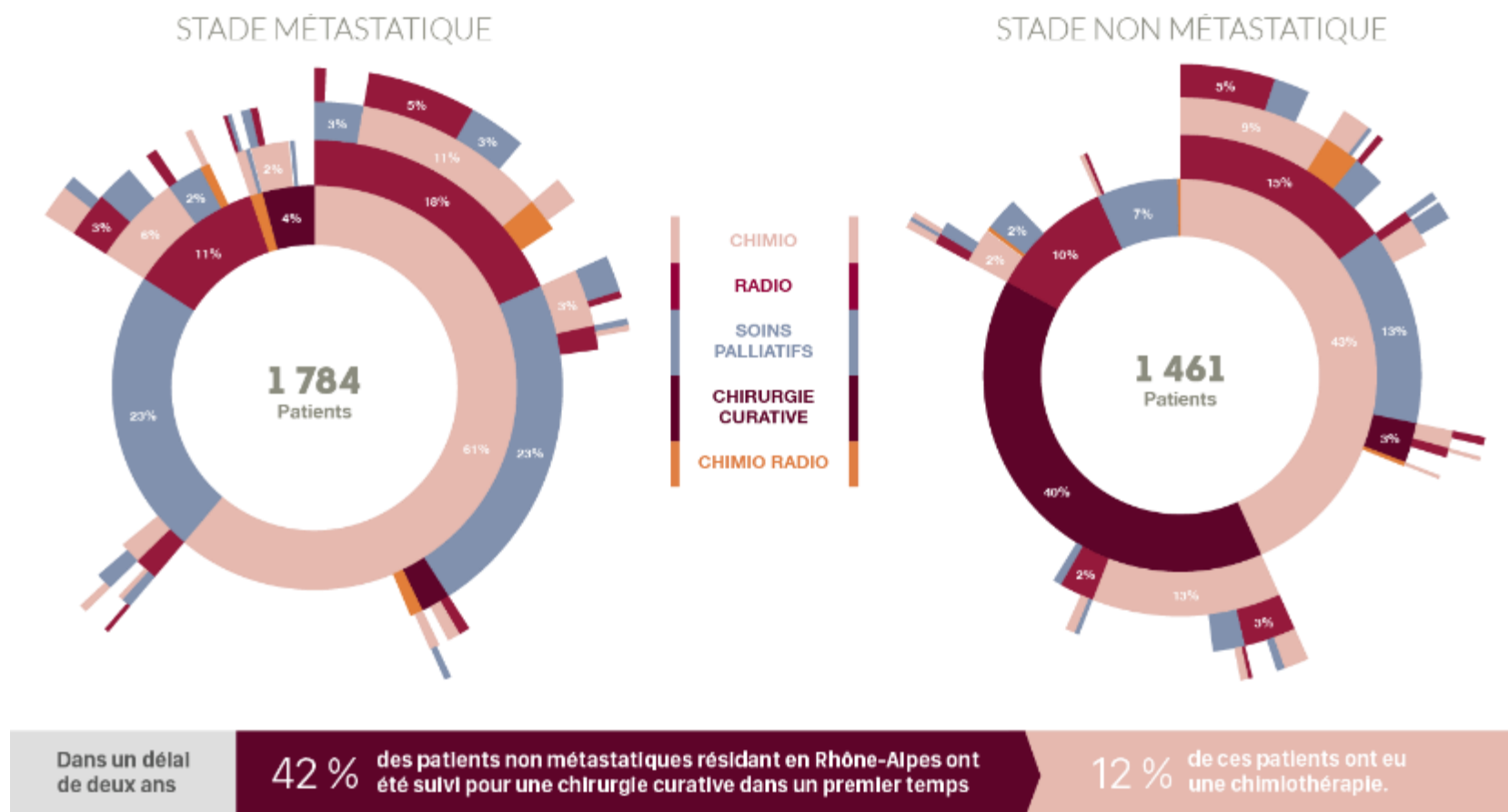
3

Analyse des Parcours de soins

Analyse des séquences de traitement direct sur 2 ans

HEVA

*patients de Rhône-Alpes
ayant eu un cancer du
poumon
Patients métastatiques et non
métastatiques*



4

Innovations méthodologiques

Du Machine learning au Process Mining

Du Data mining au machine learning



Le data mining consiste à fouiller les données qui proviennent de la collecte automatisée. Puis de les croiser avec d'autres sources de données afin de chercher une corrélation entre ces données via les différentes méthodes « classiques ».



Le passage vers le machine learning consiste à apprendre de manière automatisée et supervisée les corrélations des données afin d'être capable de produire une réponse

HEVA

“ *Le Machine Learning est un champ d'étude de l'intelligence artificielle permet ainsi d'observer **le comportement** des patients en se basant sur les faits et les traitements réels, au-delà de la description théorique du schéma thérapeutique* ”

Machine learning

*Construire des modèles
encore plus aboutis*



Classer

Classer les patients dans des groupes de profils basés sur des variables complexes et combinées



Identifier

Identifier la surexpression de variables dans un profil



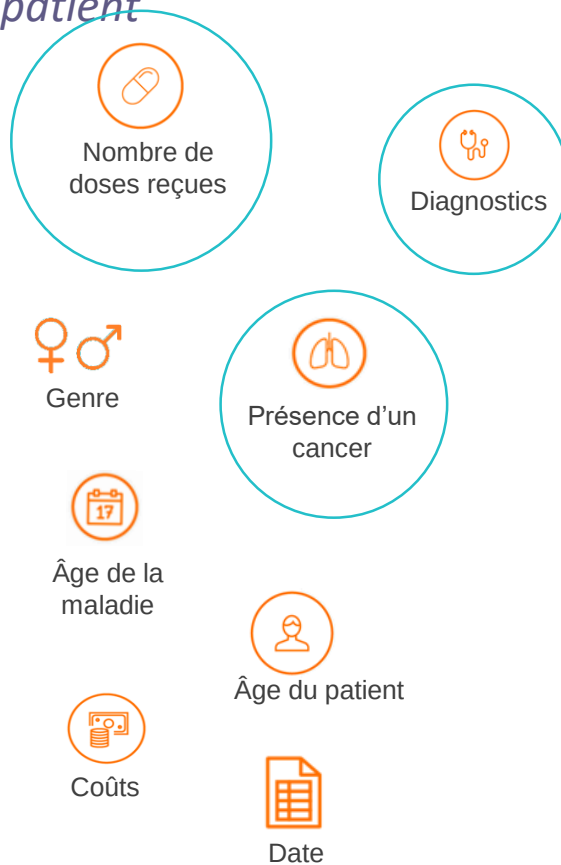
Analyser

Prédire les coûts ou les complications médicales de nouveaux patients

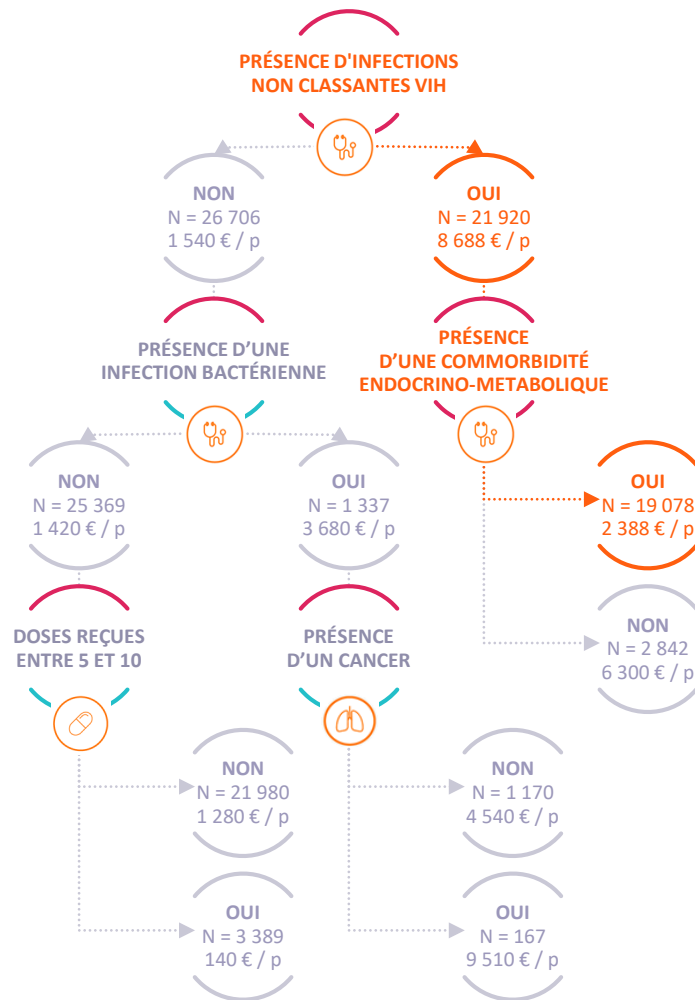
Machine learning

découvrir des profils patient

1. SÉLECTIONNER AUTOMATIQUEMENT
LES VARIABLES PERTINENTES PARMI LES
DIZAINES DISPONIBLES



2. POUR CRÉER AUTANT
DE PROFILS PATIENTS QUE BESOIN



3. EN FONCTION
D'UN OBJECTIF PRÉCIS

Expliquer le coût

ou

Mesurer des risques :
ré-hospitalisations ou présence
de complications

EXEMPLE D'UN PROFIL PATIENT

N 19 078 (63%)
PROFIL 1 COÛT MOYEN 2 388 €

CARACTÉRISTIQUES :

1. Présence d'infections non classables VIH
2. Présence d'une comorbidité endocrino-metabolique

Le machine learning

Prédire des outcomes

HEVA

1 nouveau patient

= des caractéristiques
+ un historique médical



Modèle de machine learning

« *Random Forest Classifier* »



Prédictions

85% de chances d'être réhospitalisé sous 30 jours

12% de développer une infection opportuniste

Mortalité à 12 mois accrue de 5%

Du machine learning au process mining

Discipline de recherche relativement jeune qui se trouve entre :

- l'intelligence artificielle et le Data Mining
- la modélisation et l'analyse des processus

Cette technologie se base sur les données extraites des systèmes informatiques contenant des informations sur les processus tels qu'ils se sont réellement passés



Le Process Mining une discipline qui se positionne entre l'analyse de processus et l'analyse de données qui permet ainsi d'observer la succession de prise en charge des patients et de caractériser les différents types de trajectoires



Quelles sont les finalités

1

Découverte des processus :

Découvrir, les processus tels qu'ils se sont réellement passés

2

Comparaison à un modèle de processus

Vérifier si le modèle théorique ou prévu correspond bien à la réalité

3

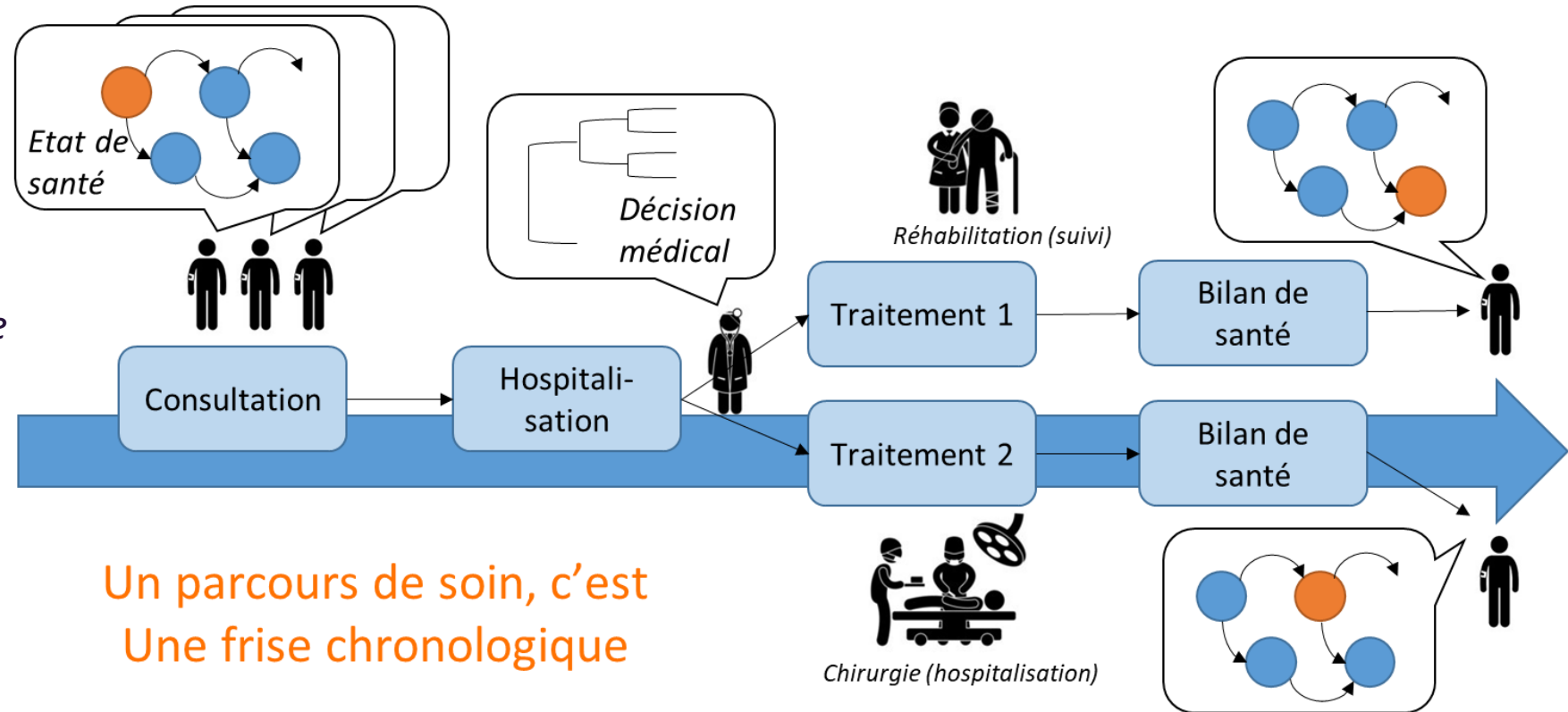
Amélioration des processus :

Mettre en place des chemins cliniques
Identifier précocement les patients à risques

Le Process Mining

l'étude des parcours

Le parcours de soins de chaque patient est unique



Le Process Mining

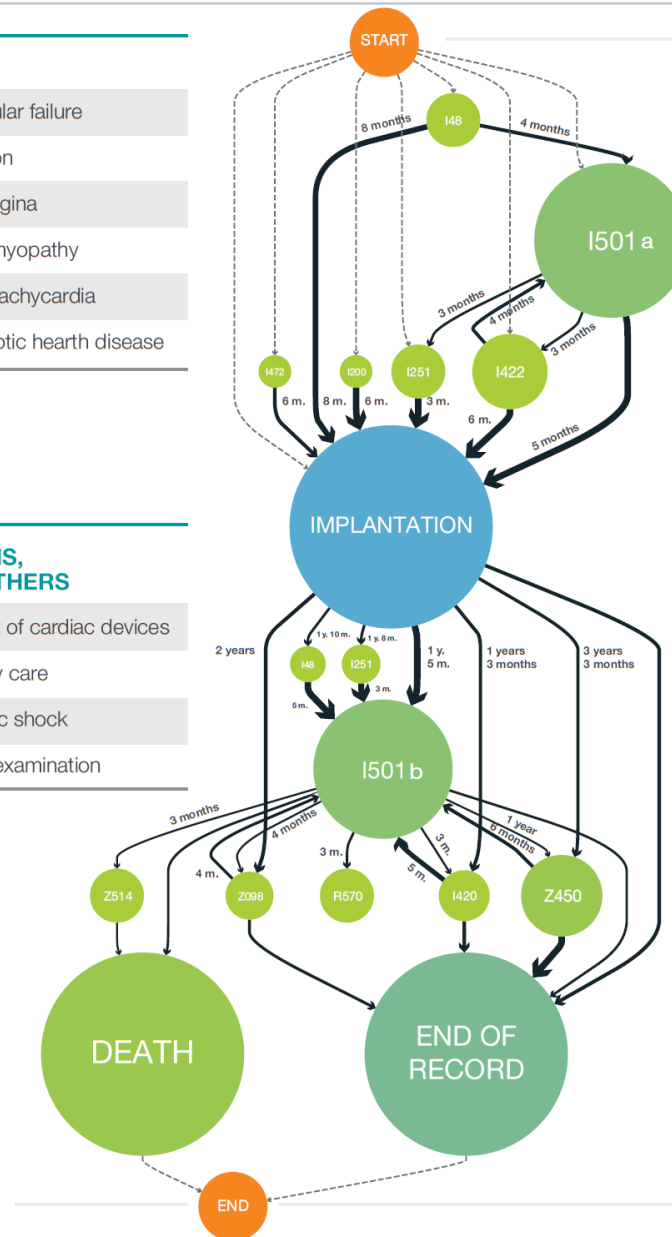
Synthèse des trajectoires en 1 modèle

HEART ISSUES

- I501 - Left ventricular failure
- I48 - Atrial fibrillation
- I200 - Unstable angina
- I42[0-2] - Cardiomyopathy
- I472 - Ventricular tachycardia
- I251 - Atherosclerotic heart disease

COMPLICATIONS, RELAPSES & OTHERS

- Z450 - Adjustment of cardiac devices
- Z514 - Preparatory care
- R570 - Cardiogenic shock
- Z098 - Follow-up examination



2006

HISTORY PRIOR TO IMPLANTATION (2006-2008)

- 51% of patients were hospitalized for heart failure (code I501), on average 5 months before implantation.
- 73% of patients had at least 1 stay in the last 6 months before implantation.
- 11% of patients did not have any history over the 2 previous years.

2008

FOLLOW-UP AFTER IMPLANTATION (2008-2013)

- 8% had complications related to their defibrillator (on average 2 years and 1 month after)
- 50% were hospitalized for heart failure (on average 8 months after), including 51% who already had heart failure before being implanted.
- 28% were readmitted to adjust their device (code Z450) (on average 2 years and 2 months after).
- 29% died at the hospital within 6 years (average : 2 years and 4 months after).

2013