

Comment tirer parti de la classification en GHM pour comparer l'activité observée durant la crise à celle qui aurait pu être attendue ?

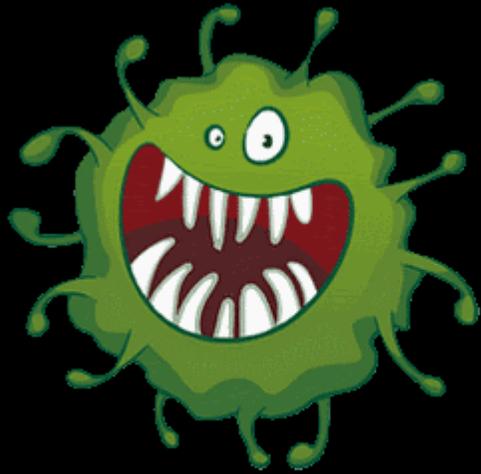
Mériem SAÏD, Marie-Caroline CLÉMENT, Joëlle DUBOIS

Service CIM-MF, ATIH

17 mars 2023

Plan

- Contexte et objectifs
- Méthodologie proposée
- Potentiel de cette méthodologie
- Bilan de la crise sanitaire
- Quelques autres cas d'usage
- Forces, limites et perspectives



Plan

- Contexte et objectifs
- Méthodologie proposée
- Potentiel de cette méthodologie
- Bilan de la crise sanitaire
- Quelques autres cas d'usage
- Forces, limites et perspectives

Contexte et objectifs

- **Contexte (1/2)**

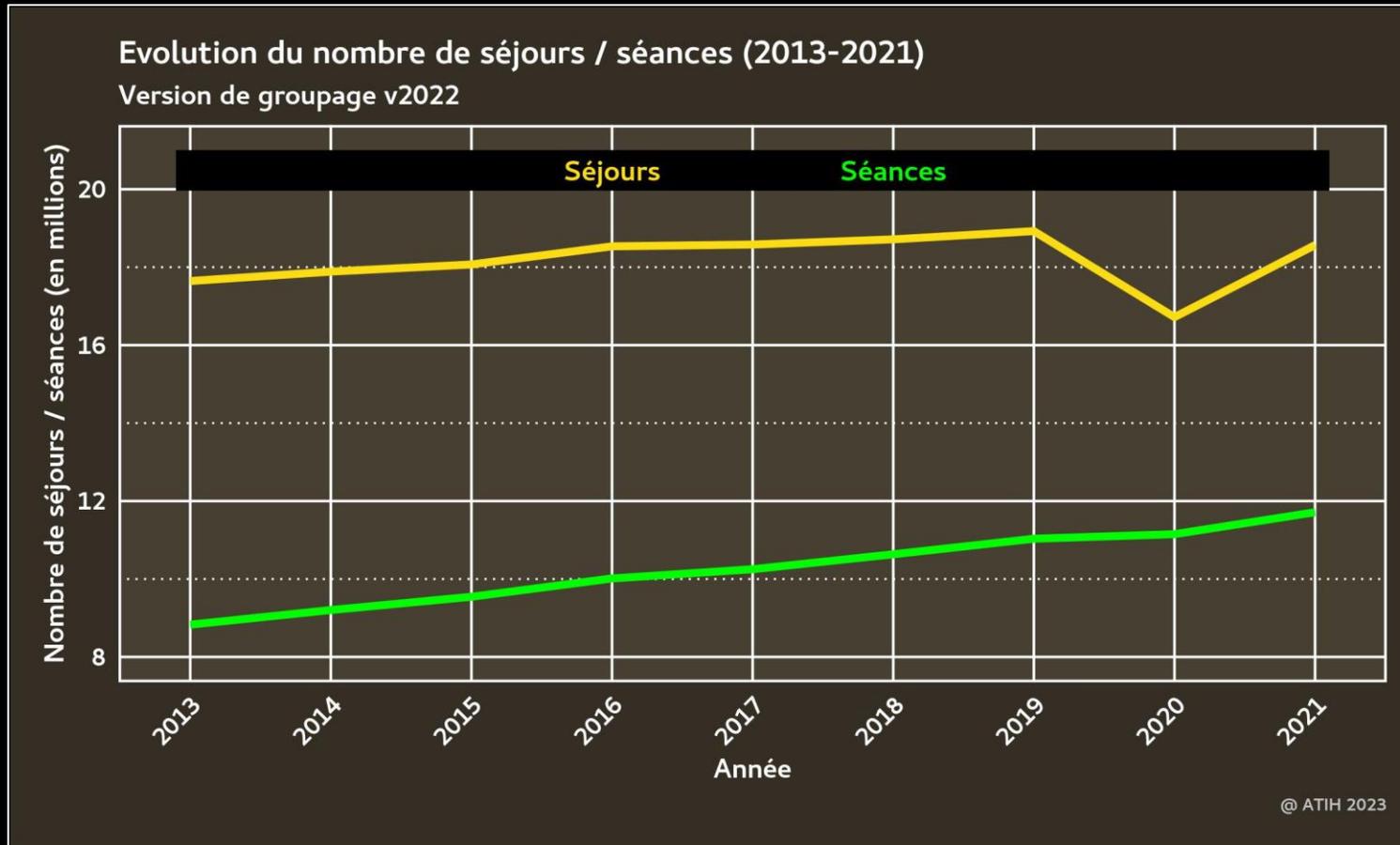
- **Mars 2020** : début de la crise sanitaire liée à la COVID-19
 - Pression sur le système hospitalier due à l'afflux de patients
 - Déprogrammations des activités non urgentes, +/- généralisées selon les vagues épidémiques
 - Effets indirects : diminution de certains facteurs de risque, « concurrence » virale, ...
- **Avant mars 2020** : dynamique loin d'être uniforme selon les activités
 - Certaines diminuaient depuis plusieurs années tandis que d'autres augmentaient ou évoluaient de façon +/- erratique

⇒ Pour effectuer un bilan de la crise, il faut prendre en compte cette **diversité d'évolution** de l'activité **durant la crise** mais également **avant la crise**

Contexte et objectifs

- **Contexte (2/2)**

- Évolution du nombre de **séjours** (hors CMD 28) et de **séances** (CMD 28)



Séjours : baisse massive en 2020 et redémarrage en 2021 sans retour à la normale

Séances : léger infléchissement en 2020 et retour apparent à la normale en 2021

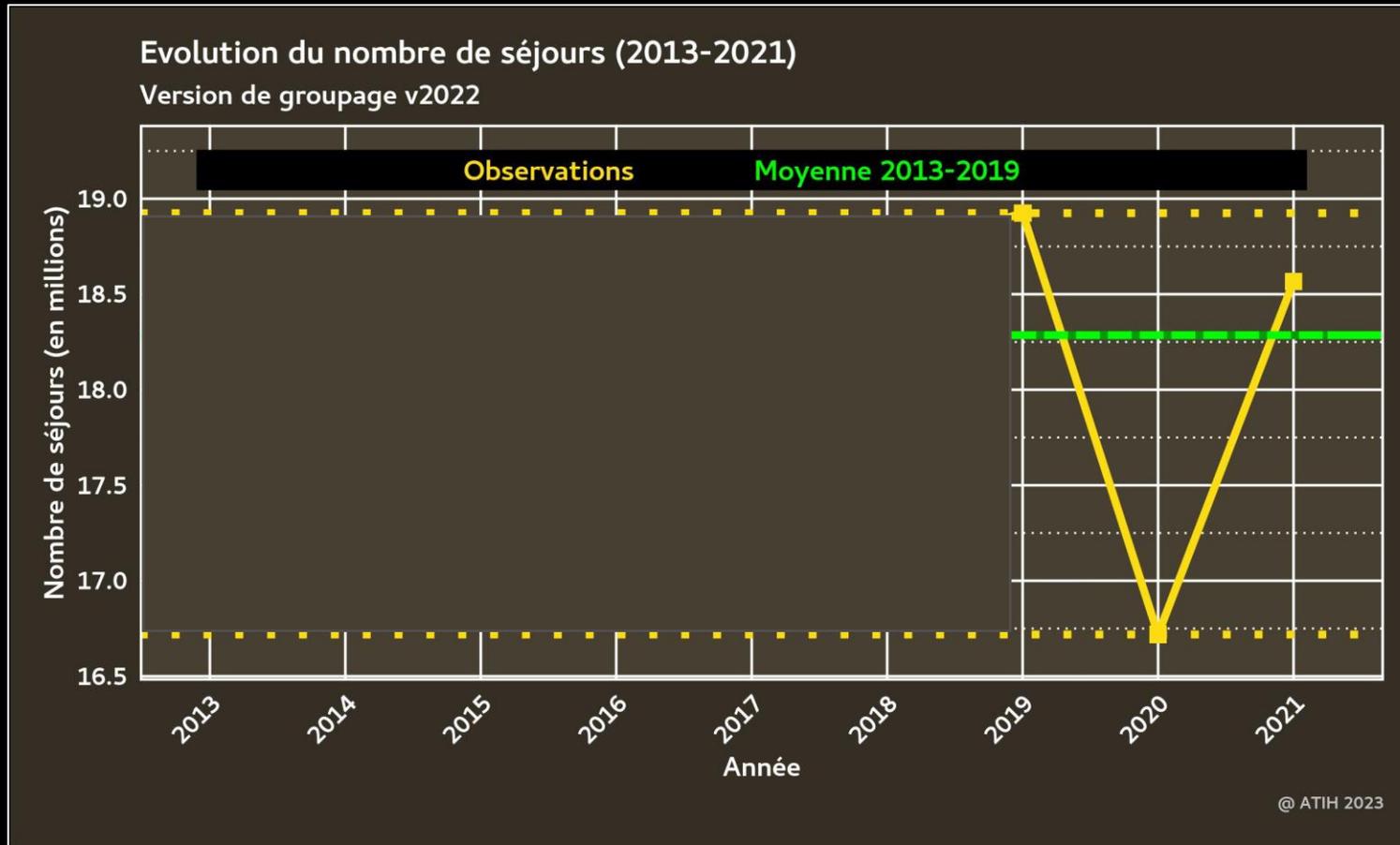
Contexte et objectifs

- **Objectifs**

- **Objectif principal** : établir un bilan aussi précis et fiable que possible de la crise sanitaire pour l'activité hospitalière en MCO en tenant compte :
 - De la diversité des impacts en fonction des pathologies et des prises en charge
 - Des évolutions antérieures à la crise sanitaire
- **Objectifs « secondaires »**
 - Repérer des évolutions « atypiques » à suivre dans les années à venir
 - Mettre en évidence d'éventuels changements de pratiques
 - Evaluer le potentiel de la méthodologie proposée à différents niveaux de granularité : région, département, établissement

Petit aparté (1/2)

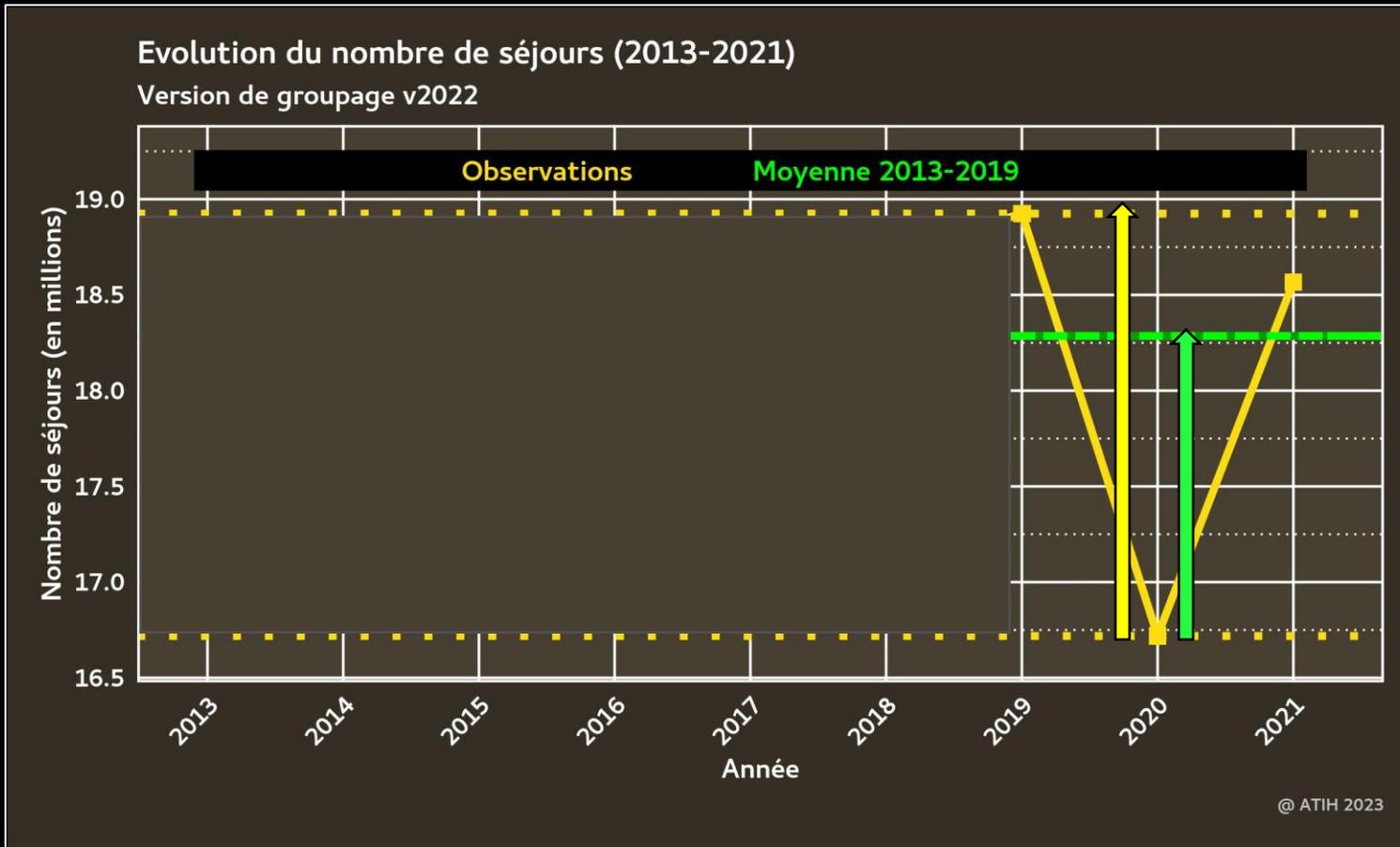
- Quel nombre de séjours aurait pu être « attendu » en 2020 par exemple ?



Nombre 2020
 ≈ 16,7 millions

Petit aparté (1/2)

- Quel nombre de séjours aurait pu être « attendu » en 2020 par exemple ?



Nombre 2019

≈ 18,9 millions

-2,2 millions

Moyenne 2013-2019

≈ 18,3 millions

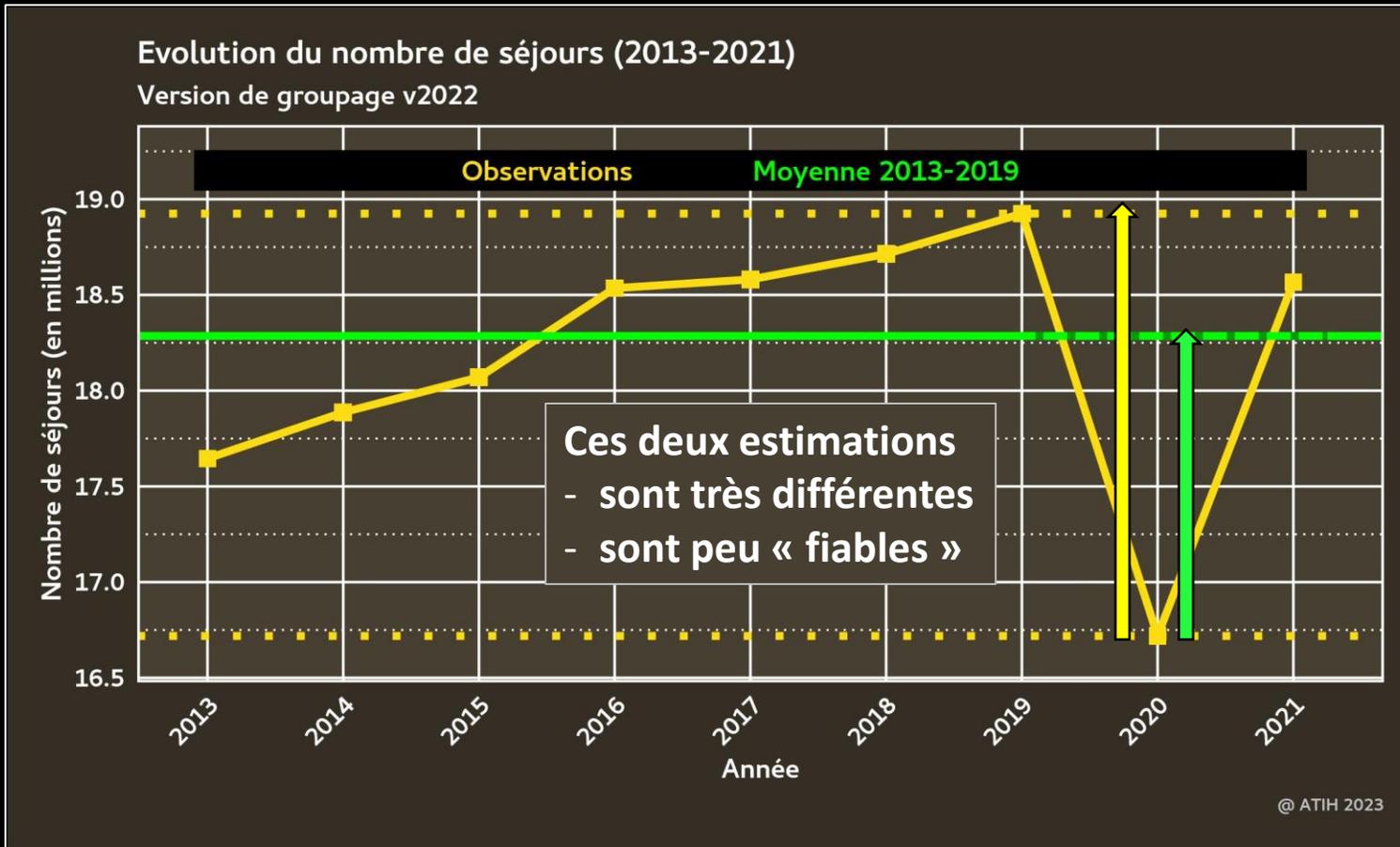
-1,6 millions

Nombre 2020

≈ 16,7 millions

Petit aparté (1/2)

- Quel nombre de séjours aurait pu être « attendu » en 2020 par exemple ?



Nombre 2019

≈ 18,9 millions

-2,2 millions

Moyenne 2013-2019

≈ 18,3 millions

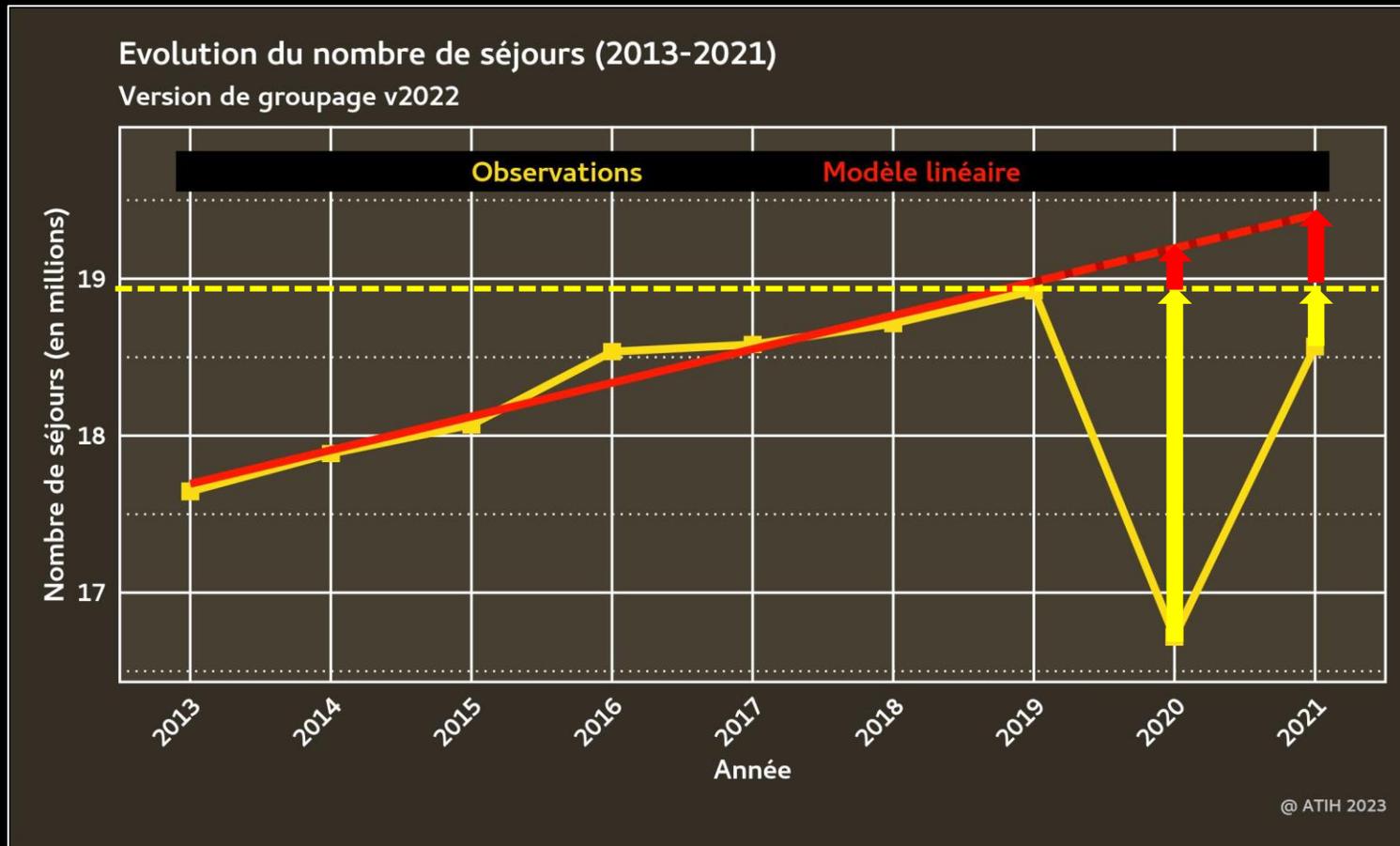
-1,6 millions

Nombre 2020

≈ 16,7 millions

Petit aparté (2/2)

- Pourquoi les 2 estimations sont-elles « erronées » ?



- Si on compare l'activité de 2020 et 2021 non pas à celle de 2019 mais à celle prédite par un **modèle linéaire**, le bilan est plus pessimiste
- **Contrainte** : dynamique pré-crise non uniforme selon les pathologies ou prises en charge
- **Solution adoptée** : modélisation pour chaque racine de GHM pour estimer le nombre de séjours qui aurait pu être attendu

Plan



- Contexte et objectifs
- Méthodologie proposée
- Potentiel de cette méthodologie
- Bilan de la crise sanitaire
- Quelques autres cas d'usage
- Forces, limites et perspectives

Méthodologie proposée

- **Données mobilisées**

- Bases PMSI MCO 2013-2021 groupées en v2022
- Exclusions : CMD 90 (erreurs), PIE prestataires et racine 27C07 (*Autres transplantations*)

- **Remarques**

- Contrainte CNIL : au plus 10 ans (9 années complètes + année en cours)
- Contrainte classification en GHM : groupage en version unique, nécessaire mais potentiellement de moins en moins fiable à mesure que l'on remonte dans le temps

Méthodologie proposée

1. Modélisation des effectifs pour chaque racine de GHM sur la période 2013-2019

- Soit N_a le nombre de séjours de l'année a pour une racine donnée
- 3 modèles sont ajustés sans autre variable explicative que le temps

$$\text{Constant : } N_a = \text{Constante} + \text{Erreur}_a$$

$$\text{Linéaire : } N_a = \text{Constante} + \text{Coeff}_1 \times a + \text{Erreur}_a$$

$$\text{Quadratique : } N_a = \text{Constante} + \text{Coeff}_1 \times a + \text{Coeff}_2 \times a^2 + \text{Erreur}_a$$

2. Choix du meilleur modèle : on favorise la parcimonie

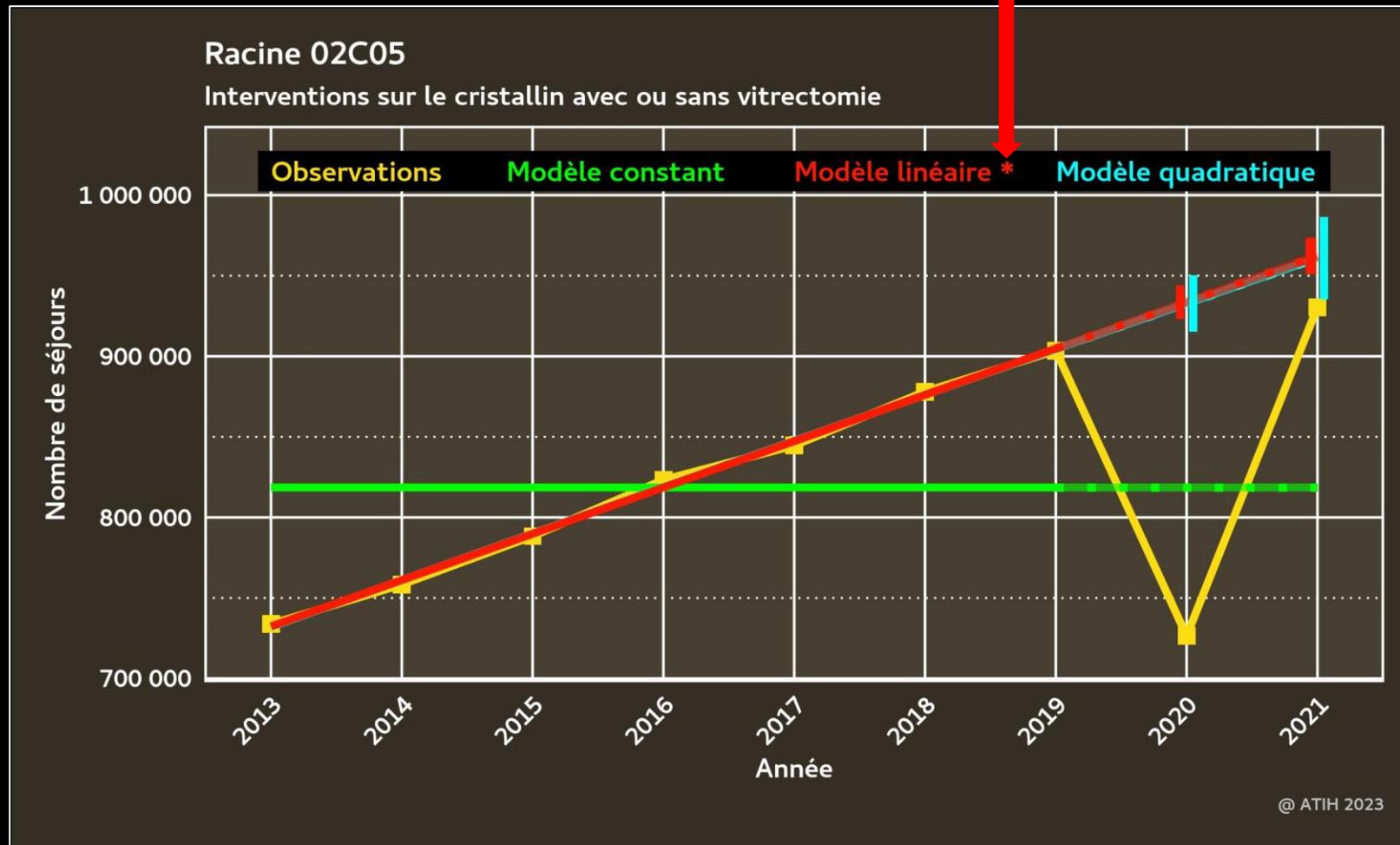
- Test statistique formel pour des modèles emboîtés
- Critère d'information d'Akaike corrigé ([pour les curieux](#))

3. Prédiction des effectifs 2020 et 2021 pour chaque racine à partir du modèle retenu

4. Comparaison prédiction / observation

Méthodologie proposée

• Exemple



Intervalles de prédiction

- 2021 toujours + large que 2020
- 2 intervalles tracés
⇔ 2 modèles significatifs ($p < 1\%$)
- A ajustements quasi-équivalents : quadratique + large que linéaire

Plan



- Contexte et objectifs
- Méthodologie proposée
- Potentiel de cette méthodologie
- Bilan de la crise sanitaire
- Quelques autres cas d'usage
- Forces, limites et perspectives

Potentiel de la méthodologie proposée

- **Poids des racines « prédictibles »**
 - Une racine est dite « **prédictible** » si un modèle linéaire ou quadratique est retenu à un seuil de significativité (*p-value*) suffisamment faible
 - « **Suffisamment faible** » \Leftrightarrow on favorise un seuil $< 1\%$, ce qui n'est pas habituel mais nécessaire eu égard au faible nombre d'observations
- **Potentiel de la méthodologie = poids des racines prédictibles**
 - Poids en racines et en séjours
 - Poids selon le modèle retenu
 - Poids selon le sens de la tendance (modèle linéaire)

Potentiel de la méthodologie proposée

• Poids des racines « prédictibles » en racines et en séjours / séances

<i>p-value</i>	Nombre de racines	Part de racines	Nombre de séjours ou séances en 2019	Part de séjours ou séances en 2019
Hors CMD 28				
<1%	375	57,3%	14 700 048	77,8%
<5%	490	74,9%	16 504 622	87,4%
Total	654	100,0%	18 889 800	100,0%
CMD 28				
<1%	17	89,5%	10 344 696	94,0%
<5%	18	94,7%	10 954 389	99,5%
Total	19	100,0%	11 009 522	100,0%

Lecture

- Hors CMD 28, 375 racines sont prédictibles avec $p\text{-value} < 1\%$, soit 57,3% des racines de séjours
- Ces 375 racines représentent environ 14,7 millions de séjours en 2019, soit 77,8% de l'ensemble des séjours de 2019

Résultats : même avec un seuil de significativité très faible (< 1%)

- Hors CMD 28 : environ 3 racines sur 5 sont prédictibles et elles couvrent plus de 3 séjours sur 4 (ou près de 4 sur 5)
- CMD 28 : près de 9 racines sur 10 sont prédictibles et elles couvrent 94% des séances

Potentiel de la méthodologie proposée

• Poids selon le type de modèle retenu

Modèle	Nombre de racines	Part de racines	Nombre de séjours ou séances en 2019	Part de séjours ou séances en 2019
Hors CMD 28, $p < 1\%$				
Linéaire	322	49,2%	13 300 744	70,4%
Quadratique	53	8,1%	1 399 304	7,4%
Constant	279	42,7%	4 189 752	22,2%
CMD 28, $p < 1\%$				
Linéaire	12	63,2%	8 986 188	81,6%
Quadratique	5	26,3%	1 358 508	12,3%
Constant	2	10,5%	664 826	6,0%

Lecture

- Hors CMD 28, 322 racines sont prédictibles par un modèle linéaire avec $p\text{-value} < 1\%$, soit 49,2% des racines de séjours
- Ces 322 racines représentent environ 13,3 millions de séjours en 2019, soit 70,4% de l'ensemble des séjours de 2019

Résultats : même avec un seuil de significativité très faible ($< 1\%$)

- Hors CMD 28 : environ 1 racine sur 2 est prédictible par un modèle linéaire, représentant plus de 7 séjours sur 10
- CMD 28 : plus de 3 racines sur 5 sont prédictibles par un modèle linéaire, représentant plus de 4 séances sur 5

Potentiel de la méthodologie proposée

- Poids selon le sens de la tendance (modèle linéaire, $p < 1\%$)

Tendance	Nombre de racines	Part de racines	Nombre de séjours ou de séances en 2019	Part de séjours ou de séances en 2019
Hors CMD 28, $p < 1\%$				
Croissante	201	62,4%	9 827 704	73,9%
Décroissante	121	37,6%	3 473 040	26,1%
Ensemble	322	100,0%	13 300 744	100,0%
CMD 28, $p < 1\%$				
Croissante	9	75,0%	8 934 959	99,4%
Décroissante	3	25,0%	51 229	0,6%
Ensemble	12	100,0%	8 986 188	100,0%

Lecture

- Hors CMD 28, parmi les 322 racines pour lesquelles un modèle linéaire est significatif à 1%, 201 ont une tendance croissante
- Ces 201 racines représentent environ 9,8 millions de séjours en 2019, soit 73,9% des séjours des racines linéairement prédictibles

Résultats : même avec un seuil de significativité très faible ($< 1\%$)

- Hors CMD 28 : plus de 3 racines sur 5 ont une tendance croissante et elles couvrent près de 3 / 4 des séjours des racines linéairement prédictibles
- CMD 28 : 3 racines sur 4 ont une tendance croissante et elles couvrent 99,4% des séances des racines linéairement prédictibles

Potentiel de la méthodologie proposée

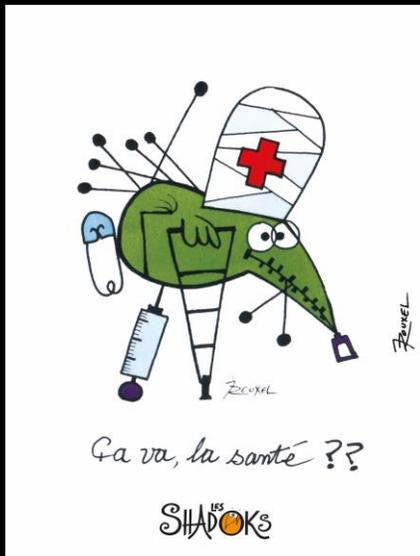
- **Principales conclusions**

La méthodologie simple qui a été proposée présente un potentiel appréciable

- Un nombre important de racines est prédictible par un simple modèle linéaire dépendant uniquement du temps, représentant plus de 70% des séjours et plus de 80% des séances de 2019 ($p\text{-value} < 1\%$)
- Ces racines présentent très majoritairement une tendance croissante
⇒ un bilan de la crise basé sur une comparaison avec la seule année 2019 présentera une sous-estimation importante

Plan

- Contexte et objectifs
- Méthodologie proposée
- Potentiel de cette méthodologie
- Bilan de la crise sanitaire
- Quelques autres cas d'usage
- Forces, limites et perspectives



Résumé de la méthodologie

Pour chacune des 673
racines de GHM

- 1 Ajustement de 3 modèles simples pour les effectifs 2013 – 2019
- 2 Choix d'1 modèle parmi les 3
- 3 Estimation des effectifs attendus pour 2020 et 2021
- 4 Calcul pour 2020 et 2021 de l'écart « observé – attendu »

Pour disposer d'un bilan
global

On additionne tous ces écarts en distinguant les racines de séjours et les racines de séances

On dispose de

- 673 valeurs d'écart pour 2020
- 673 valeurs d'écart pour 2021





Bilan de la crise sanitaire

- **Trois types de bilans ont été réalisés**
 - Ecart par rapport à 2019
 - Ecart par rapport à la moyenne de la période 2013-2019
 - Ecart par rapport à la prédiction associée au modèle retenu (linéaire, quadratique ou constant)

Bilan de la crise sanitaire

Année	Variation / 2019 Nombre	Variation / 2019 Taux	Variation / moyenne 2013–2019 Nombre	Variation / moyenne 2013–2019 Taux	Variation / modèle retenu Nombre*	Variation / modèle retenu Taux*
Hors CMD 28						
<i>* Modèle retenu avec $p < 1\%$</i>						
2020	-2 170 944	-11,5%	-1 612 861	-8,8%	-2 465 045	-12,8%
2021	-322 871	-1,7%	235 211	1,3%	-826 785	-4,3%
Total	-2 493 815		-1 377 649		-3 291 830	
CMD 28						
2020	133 825	1,2%	1 225 596	11,0%	-281 902	-2,5%
2021	687 931	6,2%	1 779 702	15,2%	-123 901	-1,1%
Total	821 756		3 005 298		-405 803	

- Variation / 2019 en nombre

Somme sur l'ensemble des racines (hors CMD 28) des écarts entre les effectifs de 2020 et les effectifs de 2019

- Variation / 2019 en taux

Somme ci-dessus divisée par les effectifs de 2019

Bilan de la crise sanitaire

Année	Variation / 2019 Nombre	Variation / 2019 Taux	Variation / moyenne 2013-2019 Nombre	Variation / moyenne 2013-2019 Taux	Variation / modèle retenu Nombre*	Variation / modèle retenu Taux*
Hors CMD 28						
<i>* Modèle retenu avec p < 1%</i>						
2020	-2 170 944	-11,5%	-1 612 861	-8,8%	-2 465 045	-12,8%
2021	-322 871	-1,7%	235 211	1,3%	-826 785	-4,3%
Total	-2 493 815		-1 377 649		-3 291 830	
CMD 28						
2020	133 825	1,2%	1 225 596	11,0%	-281 902	-2,5%
2021	687 931	6,2%	1 779 702	15,2%	-123 901	-1,1%
Total	821 756		3 005 298		-405 803	

- Variation / moyenne 2013-2019 en nombre
Somme sur l'ensemble des racines (hors CMD 28) des écarts entre les effectifs de 2020 et les effectifs moyens
- Variation / moyenne 2013-2019 en taux
Somme ci-dessus divisée par les effectifs moyens

Bilan de la crise sanitaire

Année	Variation / 2019 Nombre	Variation / 2019 Taux	Variation / moyenne 2013–2019 Nombre	Variation / moyenne 2013–2019 Taux	Variation / modèle retenu Nombre*	Variation / modèle retenu Taux*
Hors CMD 28						
<i>* Modèle retenu avec $p < 1\%$</i>						
2020	-2 170 944	-11,5%	-1 612 861	-8,8%	-2 465 045	-12,8%
2021	-322 871	-1,7%	235 211	1,3%	-826 785	-4,3%
Total	-2 493 815		-1 377 649		-3 291 830	
CMD 28						
2020	133 825	1,2%	1 225 596	11,0%	-281 902	-2,5%
2021	687 931	6,2%	1 779 702	15,2%	-123 901	-1,1%
Total	821 756		3 005 298		-405 803	

- Variation / modèle retenu en nombre

Somme sur l'ensemble des racines (hors CMD 28) des écarts entre les effectifs de 2020 et les effectifs prédits pour 2020

- Variation / modèle retenu en taux

Somme ci-dessus divisée par les effectifs prédits pour 2020

Bilan de la crise sanitaire

Année	Variation / 2019 Nombre	Variation / 2019 Taux	Variation / moyenne 2013-2019 Nombre	Variation / moyenne 2013-2019 Taux	Variation / modèle retenu Nombre*	Variation / modèle retenu Taux*
Hors CMD 28						
* Modèle retenu avec $p < 1\%$						
2020	-2 170 944	-11,5%	-1 612 861	-8,8%	-2 465 045	-12,8%
2021	-322 871	-1,7%	235 211	1,3%	-826 785	-4,3%
Total	-2 493 815		-1 377 649		-3 291 830	
CMD 28						
<div style="border: 1px solid black; padding: 5px; display: inline-block;"> ≈ 800 000 séjours d'écart </div>						
2020	133 825	1,2%	1 225 596	11,0%	-281 902	-2,5%
2021	687 931	6,2%	1 779 702	15,2%	-123 901	-1,1%
Total	821 756		3 005 298		-405 803	
<div style="border: 1px solid black; padding: 5px; display: inline-block;"> ≈ 1 230 000 séances d'écart </div>						

- Utiliser la moyenne de la période 2013-2019, ce qui revient à utiliser un modèle constant pour toutes les racines, est une très mauvaise idée.
- Cela n'est pas étonnant puisque les résultats ont montré qu'un grand nombre de racines présentent une tendance linéaire.
- Pour cette même raison et le fait que la majorité des tendances significatives sont croissantes, le bilan établi par rapport au modèle retenu est plus pessimiste mais aussi plus fiable que par rapport à la seule année 2019.
- La différence entre les deux est bien loin d'être anodine...

Plan

- Contexte et objectifs
- Méthodologie proposée
- Potentiel de cette méthodologie
- Bilan de la crise sanitaire
- Quelques autres cas d'usage
- Forces, limites et perspectives



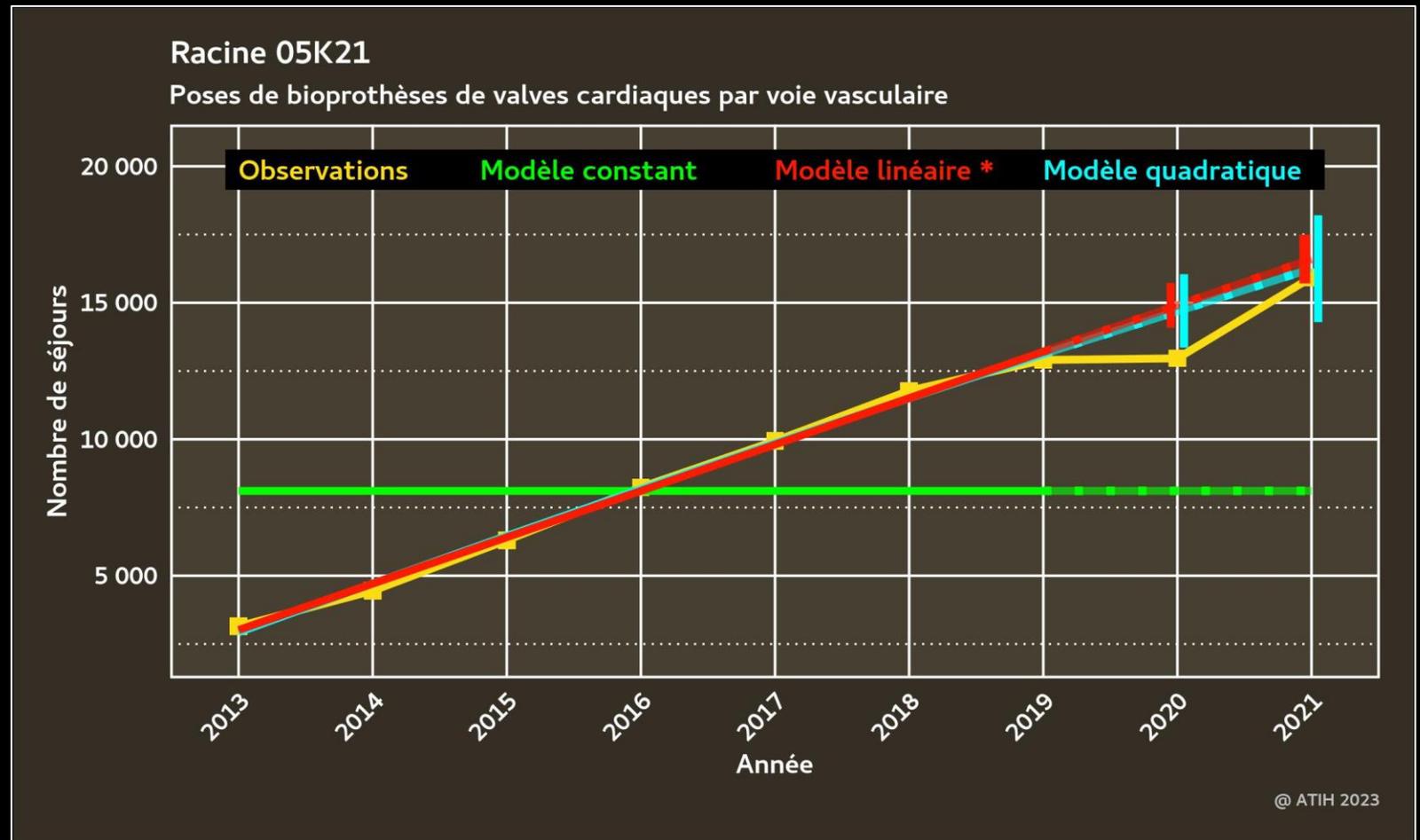
Quelques autres cas d'usage

- **Les prédictions d'activité par racine de GHM présentent également une utilité pour :**
 - 1. Suivre l'activité hospitalière et identifier des évolutions atypiques ou inattendues**, c'est-à-dire non expliquées par des évolutions du recueil PMSI, de la classification médico-économique ou des modalités de financement
 - Documenter les évolutions d'activité (suivi, veille sanitaire)
 - Émettre des hypothèses médicales pouvant expliquer ces évolutions (évolutions de pratique, épidémies, *etc.*)
 - Informer / alerter les institutions compétentes qui pourront mener des investigations si elles les jugent nécessaires
 - 2. Analyser et comparer les évolutions de l'activité à des niveaux plus fins** (région, département ou établissement)
 - Organisation de l'offre de soins
 - Pilotage des établissements de santé, *etc.*

Quelques autres cas d'usage

Suivre l'activité hospitalière et identifier des évolutions atypiques ou inattendues

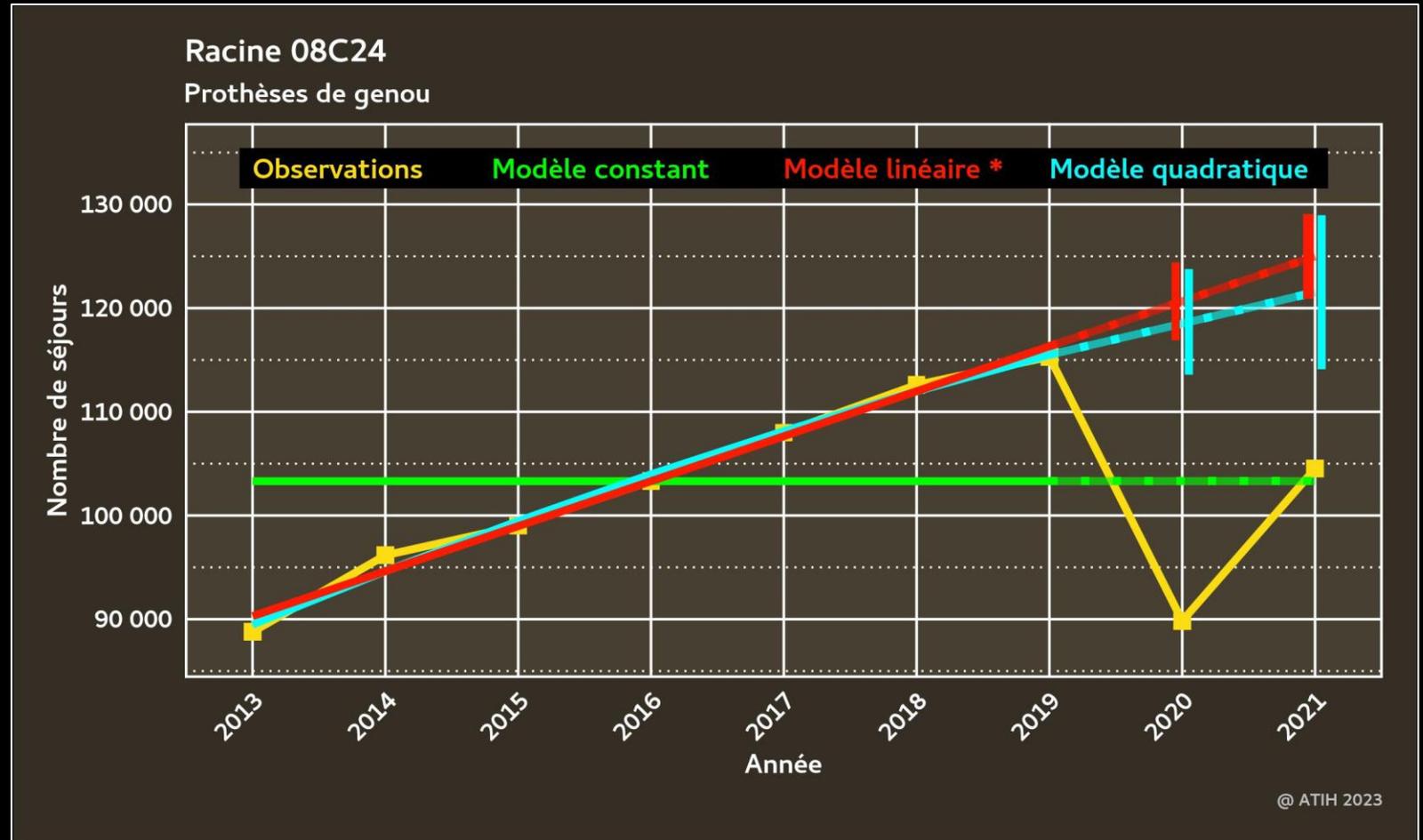
- L'activité de pose de bioprothèses de valves cardiaques par voie vasculaire semblait s'être maintenue en 2020 (niveau activité \approx 2019)
- Mais forte tendance à la hausse depuis 2013
 \Rightarrow impact de la crise sanitaire en 2020 visible
- Activité 2021 \approx activité qui aurait été attendue par rapport à l'évolution 2013-2019 : reprise et retour au niveau d'activité attendu mais pas de rattrapage de l'activité non réalisée en 2020



Quelques autres cas d'usage

Suivre l'activité hospitalière et identifier des évolutions atypiques ou inattendues

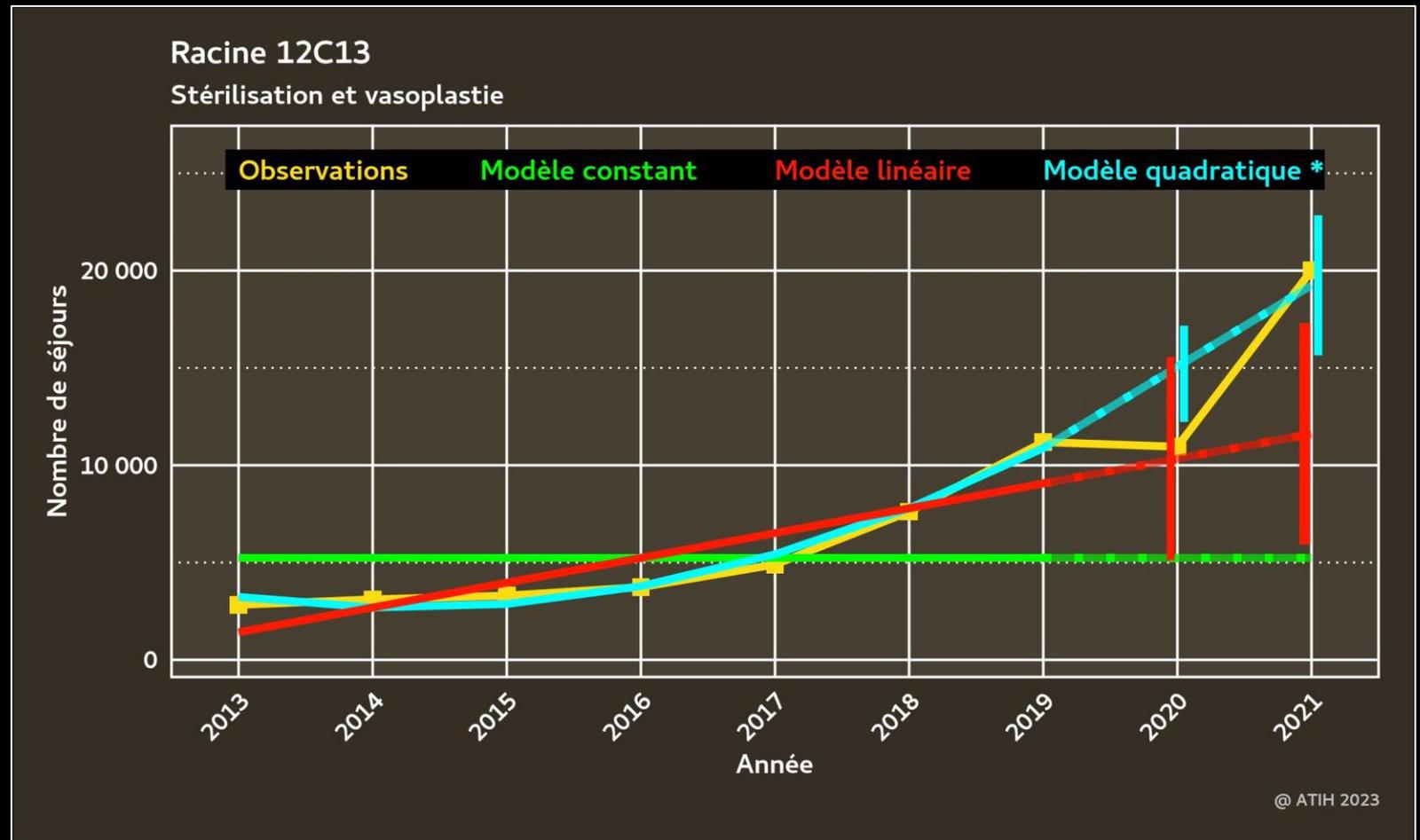
- L'activité de prothèses de genou a subi une très forte baisse en 2020
- Activité 2021 toujours très en baisse par rapport à l'activité 2019 et à l'activité attendue 2021 par rapport à l'évolution 2013-2019
- Impact de la crise sanitaire toujours visible en 2021



Quelques autres cas d'usage

Suivre l'activité hospitalière et identifier des évolutions atypiques ou inattendues

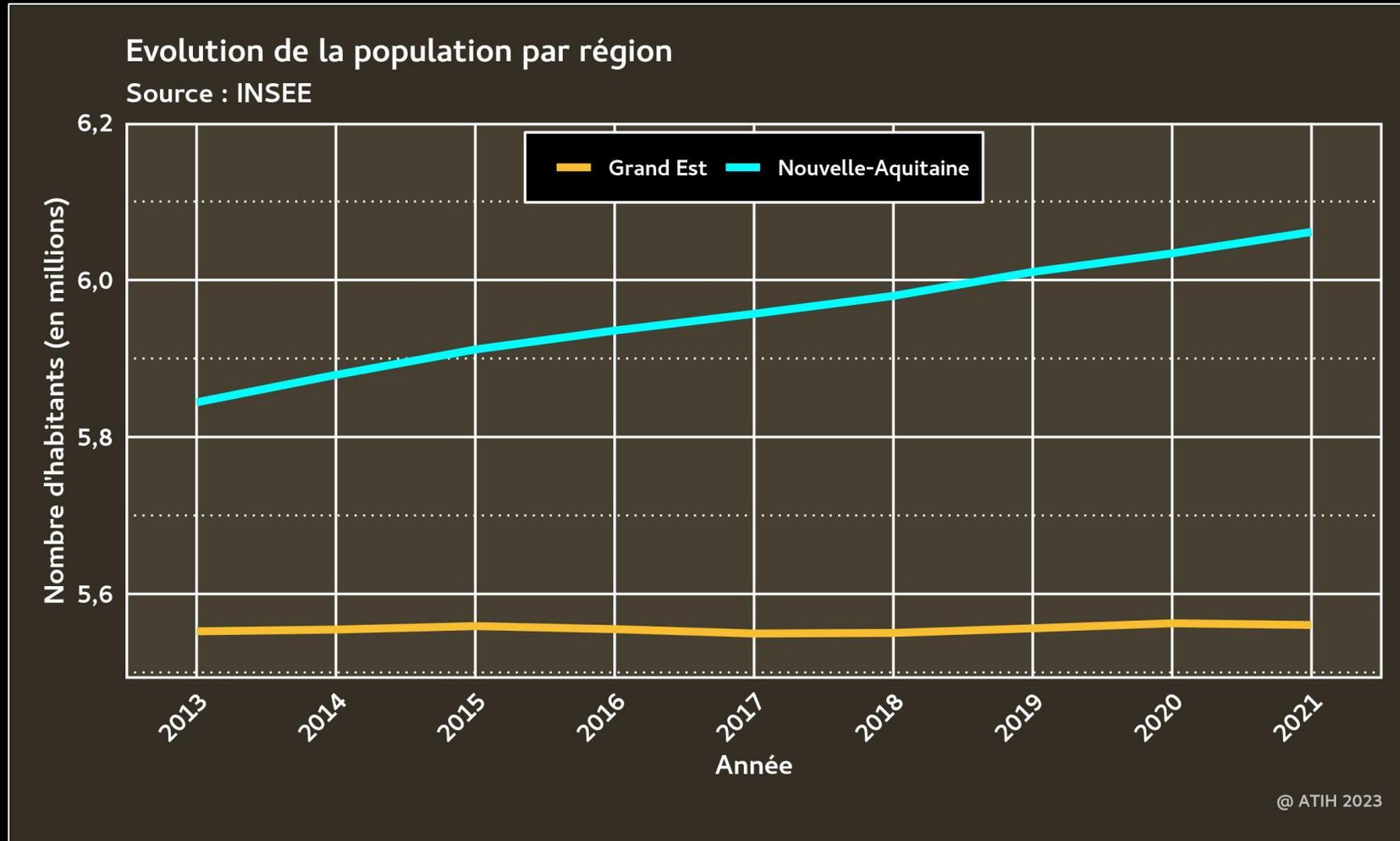
- Une des rares activités qui semblait maintenue en 2020 (niveau \approx 2019) et très forte augmentation de l'activité en 2021
- Mais, si on prend en compte l'évolution 2013-2019, l'interprétation change : impact fort de la crise sanitaire en 2020 et retour au niveau d'activité attendu en 2021



Quelques autres cas d'usage

Analyser et comparer les évolutions de l'activité aux niveaux région, département, étab.

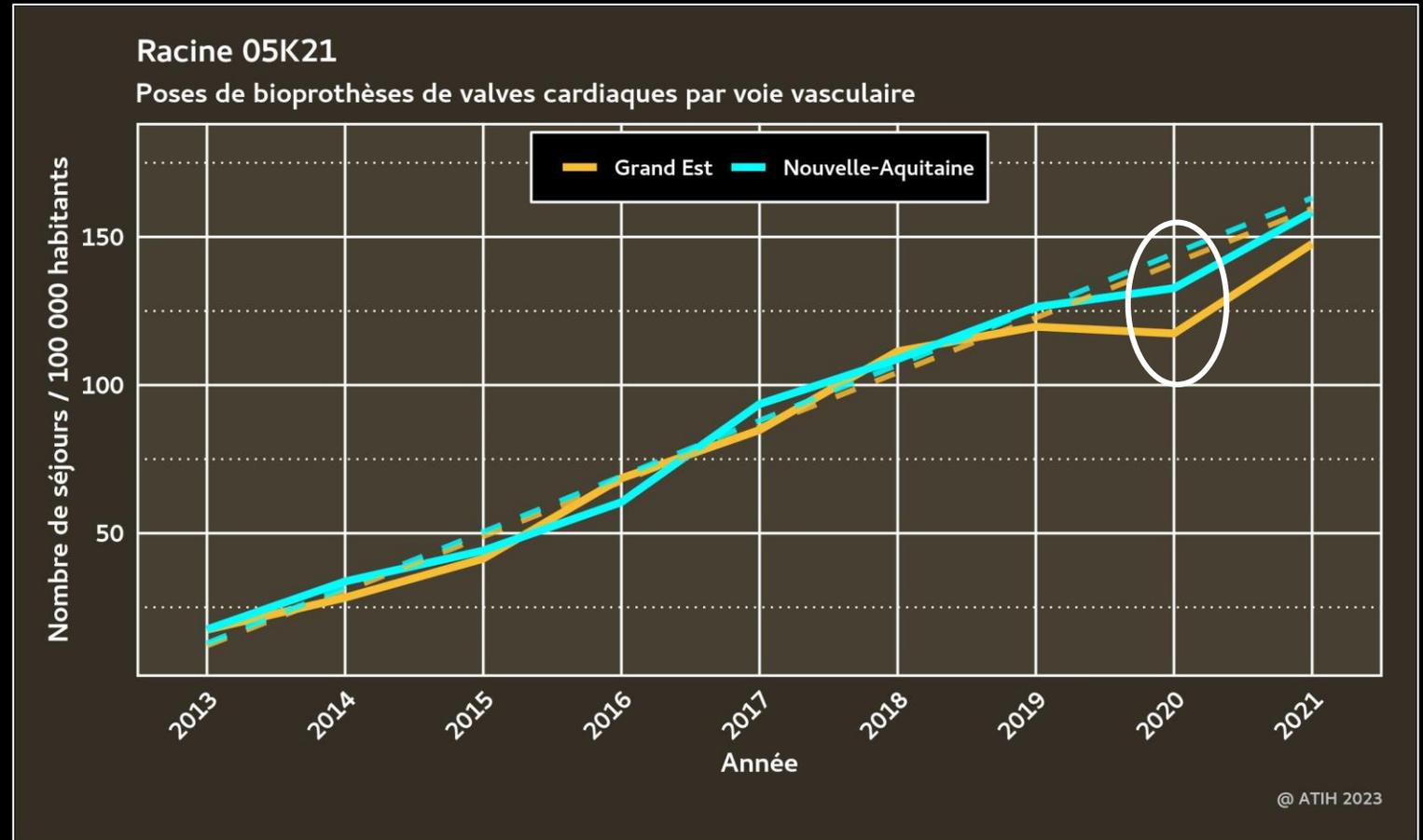
- Comparaison de deux régions pour quelques racines



Quelques autres cas d'usage

Analyser et comparer les évolutions de l'activité aux niveaux région, département, étab.

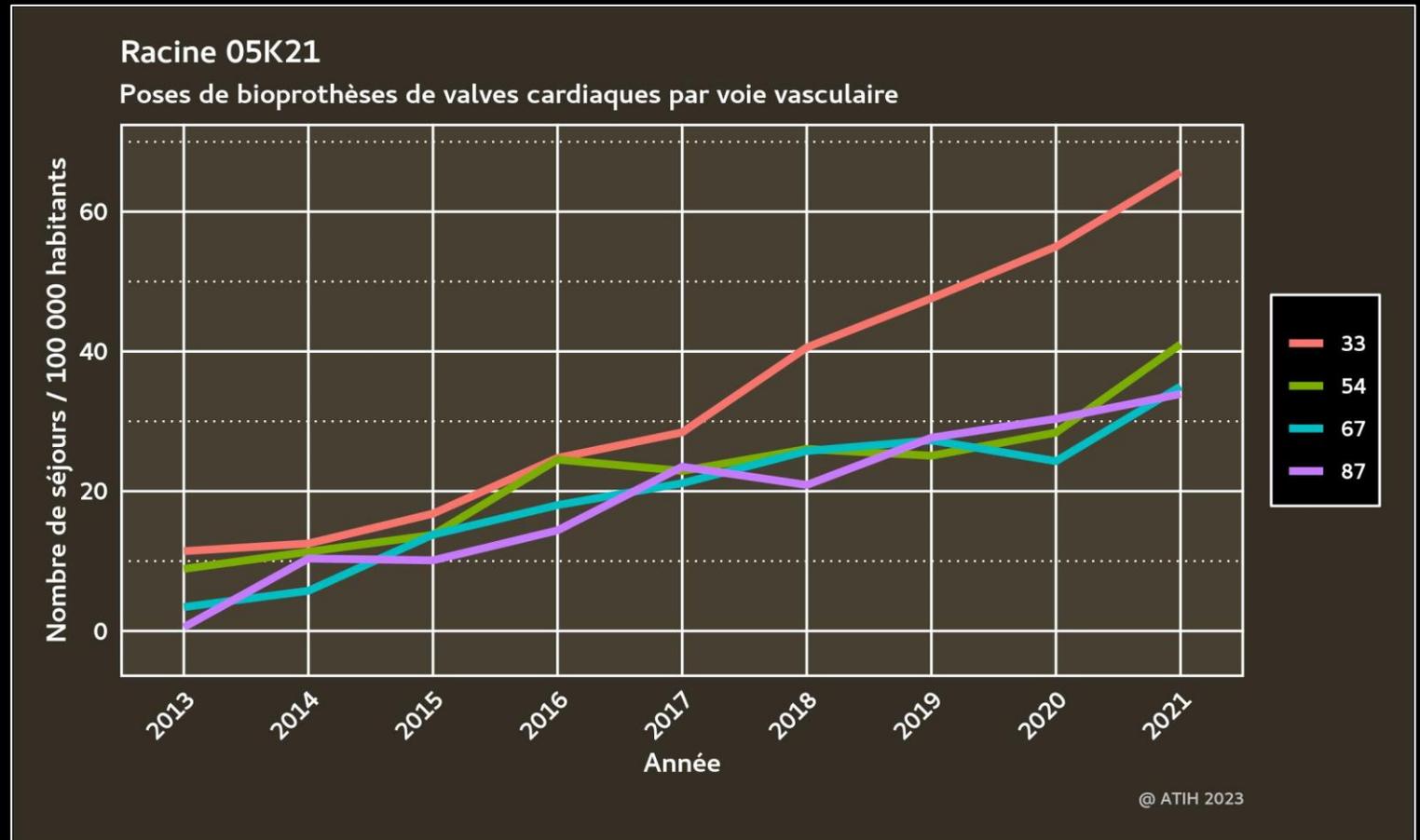
- L'activité de pose de bioprothèses de valves cardiaques par voie vasculaire a une évolution de l'activité 2013-2019 proche entre les régions Grand Est (GE) et Nouvelle-Aquitaine (NA)
- L'impact de la crise sanitaire en 2020 est plus visible pour la région GE que NA ← la région GE a été plus touchée par la Covid



Quelques autres cas d'usage

Analyser et comparer les évolutions de l'activité aux niveaux région, département, étab.

- Quand on analyse l'activité de la racine 05K21 par département, l'impact est variable :
 - Entre les départements d'une même région (67 versus 54)
 - Entre les départements de régions différentes (33 versus 67)

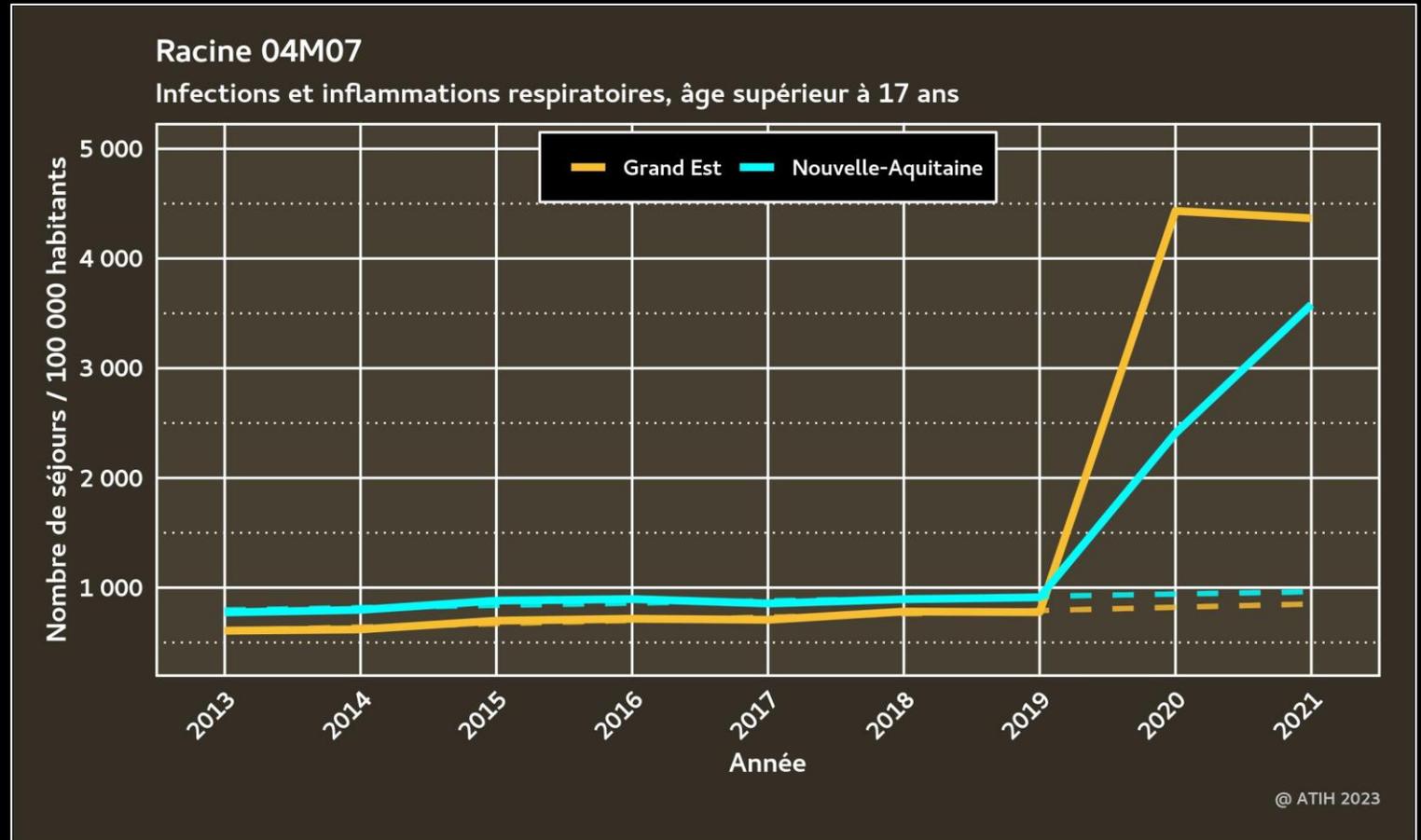


La figure représente pour chaque région le département avec l'activité la plus importante (i) en nb de séjours (67 & 33) et (ii) en nb de séjours/100 000 habitants (54 & 87).

Quelques autres cas d'usage

Analyser et comparer les évolutions de l'activité aux niveaux région, département, étab.

- 04M07 = racine de groupage des séjours Covid des ≥ 18 ans
- L'impact de la crise sanitaire en 2020 est plus visible pour la région GE que NA \Leftarrow la région GE a été plus touchée par la Covid



Quelques autres cas d'usage

Analyser et comparer les évolutions de l'activité aux niveaux région, département, étab.

- Évolution proche des départements de chaque région
- Mais un nb de séjours / 100 000 hab. inférieur en NA qu'en GE

Racine 04M07

Infections et inflammations respiratoires, âge supérieur à 17 ans

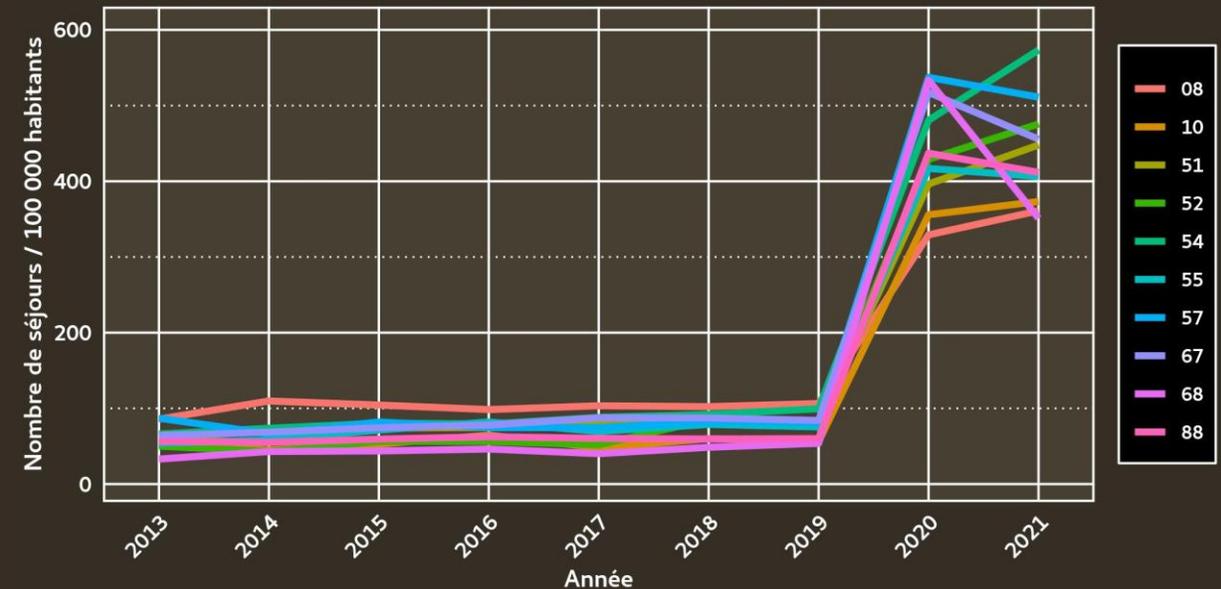
Nouvelle Aquitaine



Racine 04M07

Infections et inflammations respiratoires, âge supérieur à 17 ans

Grand Est



@ ATIH 2023

Plan

- Contexte et objectifs
- Méthodologie proposée
- Potentiel de cette méthodologie
- Bilan de la crise sanitaire
- Quelques autres cas d'usage
- Forces, limites et perspectives



Forces, limites et perspectives

- **Forces**

- Méthodologie simple et aisément reproductible
- Extensions possibles à différents niveaux de granularité

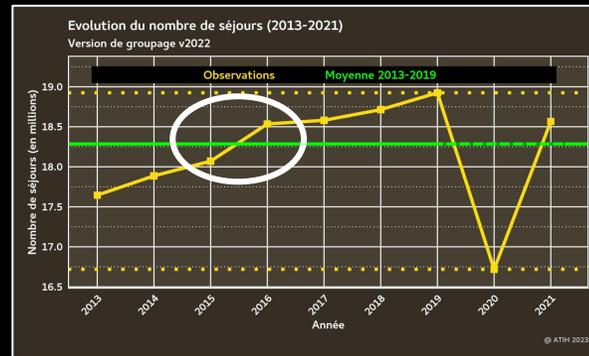
- **Limites**

- Très faible nombre d'observations historiques \Rightarrow peu de marge de manœuvre pour enrichir les modèles avec d'autres variables explicatives

Forces, limites et perspectives

- Perspectives

- Tester la prédictibilité par racine ou sous-CMD pour différents niveaux de granularité : établissement, catégorie d'établissements, région, ... (en cours)
- Affiner la modélisation pour tenir compte des modifications des règles de recueil des « séjours contigus »



- Enrichir les données en utilisant les observations mensuelles

