

## Introduction

Estimer la prévalence de la Maladie Rénale Chronique (MRC) Non Dialysée (ND) est un enjeu majeur de santé publique, et pourtant cette donnée n'est pas disponible actuellement sur la population Française.

## Objectif

Identifier l'ensemble des patients avec une insuffisance rénale chronique non dialysée en France, codée ou non via un diagnostic CIM-10 dans une base de remboursement médico-administrative, en utilisant un algorithme de Machine Learning (ML).



Un grand nombre de patients MRC ne sont pas identifiables via des codes ou des proxy.

## Source de données

Étude de cohorte rétrospective conduite à partir du Système National des Données de Santé (SNDS), en particulier à partir de l'EGB (Echantillon National des Bénéficiaires), qui est un échantillon vivant représentatif du SNIIRAM au 1/97ème.



Les patients des différentes bases de données sont chaînés au moyen d'un identifiant pseudonyme unique calculé et attribué à chaque individu.

## Design de l'étude

### Cohorte 1 = Patients MRC confirmés via un diagnostic codé dans la base

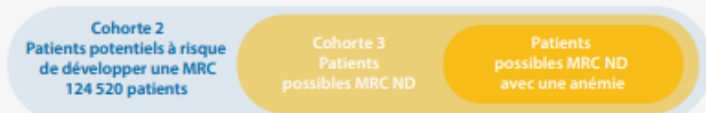
Inclusion des patients adultes ( $\geq 18$  ans) avec une hospitalisation ou une ALD en lien avec la MRC (catégorie CIM-10 N18) entre le 01/01/2012 et le 31/12/2017.



### Cohorte 2 = Patients avec une pathologie cardiovasculaire, une hypertension artérielle ou un diabète et potentiellement plus à risque de développer une MRC

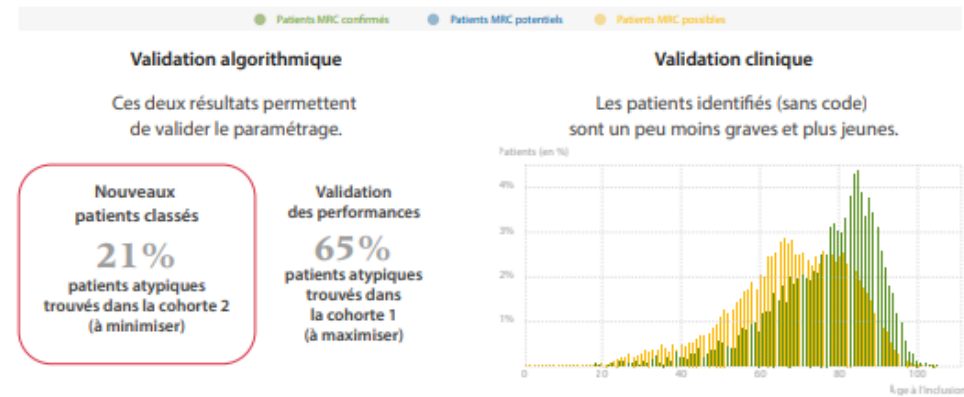
Inclusion des patients adultes ( $\geq 18$  ans) avec une ALD ou un traitement en lien avec un diabète, une hypertension artérielle ou une pathologie cardiovasculaire<sup>1</sup> entre le 01/01/2012 et le 31/12/2017.

Identification des patients possibles MRC via un algorithme de ML.

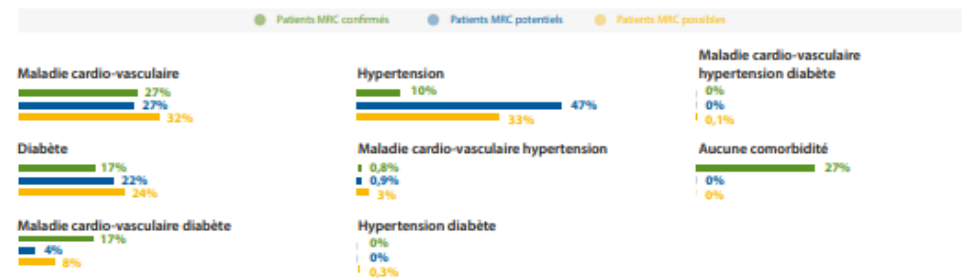


## DAKOTAH : L'intelligence artificielle au service de l'épidémiologie : Estimation du nombre de patients avec une insuffisance rénale chronique non dialysée en France

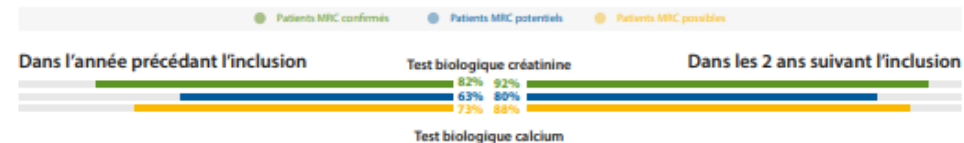
### Population trouvée par l'algorithme



### Comorbidités à l'inclusion



### Présence de traitements et tests biologiques



## Méthodes d'analyses

- Sélection de la cohorte 2** : patients potentiels à risque de MRC
- Recherche des patients atypiques au sein de la cohorte 2** : application d'un algorithme de ML [One Class SVM] afin de déterminer, parmi les patients de la cohorte 2, ceux qui sont «possibles MRC» sans présence de diagnostic CIM-10 codé dans la base de données.
- Validation grâce aux patients confirmés** : cet algorithme sépare les patients atypiques (MRC) des autres à partir de variables jugées pertinentes, puis valide la pertinence de cette séparation à partir de la cohorte 1 (patients codés MRC).



● Patients MRC confirmés ● Patients MRC potentiels ● Patients MRC possibles

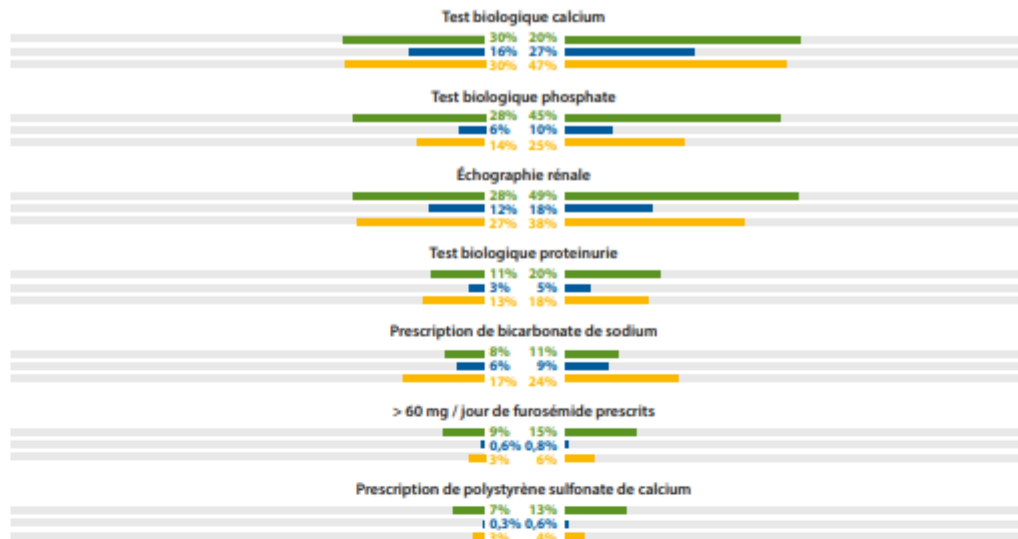
## Variables retenues dans l'algorithme de ML

- Sexe
- Nombre / durée des hospitalisations pour un motif rénal
- Nombre de consultations chez le médecin généraliste
- Traitements : Lasilix® (furosémide>80mg), Resikali®, Kayekalate®, Renvela®, Renagel®, Fosrenol®, Velporo®, Phosphosorb®, gélule de Bicarbonate de calcium, Unalfa®, Rocaltrol®
- Examens biologiques : créatinine, protéinurie, calcium, phosphate, PTH, biopsie rénale, échographie rénale

## Conclusion

Cette étude a permis d'estimer 2,3 millions de patients MRC ND en France dont 1,8 million non diagnostiqués.

Leur identification en routine permettrait une prise en charge précoce, ce qui aurait un impact potentiel en termes de santé publique.



Les patients choisis par l'algorithme ressemblent aux patients connus, notamment au niveau des traitements et tests spécifiques aux patients MRC

## Résultats étude DAKOTAH prévalence en 2017 en population générale

MRC ND		Anémiques et MRC ND	
Population générale	Algorithme ML	Population générale	Algorithme ML
0,88	4,22	Taux pour 100 personnes	0,38
470 503*	2 556 274*	Nombre de patients	140 534**
			473 377**

## Validation par la littérature

473 102 MRC stades 3-5 suivis par un néphrologue <sup>2</sup>	132 417 MRC suivis par un néphrologue Hb < 12g/dL <sup>3</sup>
--	---

\*Prévalence X 53,47 million (population adulte Française 2020 INSEE \*)

\*\*Prévalence X 53,47 million (population adulte Française 2020 INSEE \*) X 69,17% (Stades 3-5 dans l'étude DAKOTAH)



## Remerciements

Cette étude a été financée par Astellas. Les auteurs remercient Gwendoline Poincot et Julien Beisel pour la réalisation des analyses et Mélanie Bouillard pour la conception du poster.



## Transparence

Cette étude a été approuvée par le Health Data Hub le 7 avril 2020 (dossier n°778539).

## Références

- Cartographie des pathologies et des dépenses de l'Assurance Maladie. Disponible en ligne : <https://assurance-maladie.ameli.fr/etudes-et-donnees/par-theme/pathologies/cartographie-assurance-maladie>
- Hill, N. R., Fatoba, S. T., Oka, J. L., Hirt, J. A., O'Callaghan, C. A., Lasserson, D. S., and Hobbs, F. D. R. (2016). Global Prevalence of Chronic Kidney Disease – A Systematic Review and Meta-Analysis. PLOS ONE, 11(7), e0158765.

- Wong, M. (2020). Anemia and iron deficiency among chronic kidney disease Stages 3-5ND patients in the Chronic Kidney Disease Outcomes and Practice Patterns Study: often unmeasured, variably treated | Clinical Kidney Journal | Oxford Academic.
- INSEE. <https://www.insee.fr>

## Abréviations

ALD : Affection Longue Durée  
CIM : Classification Internationale des Maladies  
DCRI : Datamart de Consommation Inter Régime  
EGB : Échantillon National des Bénéficiaires  
ML : Machine Learning  
MRC : Maladie Rénale Chronique  
NIR : Numéro d'Inscription au Répertoire

ND : Non Dialysée  
PMS : Programme de médicalisation des systèmes d'information  
SNDIS : Système National des Données de Santé  
SNIRAM : Système national d'information inter-régimes de l'Assurance maladie  
SVM : Support Vector Machine