



Les grandes banques de données médicales et administratives du Québec pour la recherche en santé

Peut-on innover?



Université de
Sherbrooke

Yohann M Chiu, PhD

Département de médecine de
famille et médecine d'urgence

5 juin 2024
Présentation à l'EINS 2024



Conflits d'intérêt

- Postdoc: IRSC, CRSNG, FRQS et INSPQ
- Travaux actuels: CRCHUS, IRSC
- Aucun autre conflit à signaler



- I. Les banques de données médico-administratives pour la recherche
 - Avantages et inconvénients
 - SISMACQ
- II. Grande utilisation des services d'urgence
 - Profils des grands utilisateurs
 - Construction d'un outil de dépistage
- III. Surveillance de la polypharmacie
 - Fouille de données
 - Prédiction
 - Transparence
- IV. Conclusions et réflexions
 - On n'a pas parlé de...
 - Statistiques et IA pour la recherche en santé

I – Avantages et inconvénients

Utilisation secondaire possible des données médico-administratives

- + Design alternatif pour mesurer l'utilisation, l'efficacité et la sécurité
- + Évaluer les traitements efficaces dans des **contextes réels** (études observationnelles)
- - Moins de contrôle sur les conditions d'évaluation de la recherche
- - Banques de données **administratives**



Using Secondary Data Sources for Pharmacoepidemiology and Outcomes Research

Spencer E. Harpe Pharm.D., Ph.D., M.P.H.

First published: 06 January 2012 | <https://doi.org/10.1592/phco.29.2.138> | Citations: 54

PDF TOOLS SHARE



Canadian Journal of Cardiology

Volume 28, Issue 2, March–April 2012, Pages 162-168



Clinical research - health care outcome

Evaluation of Care and Surveillance of Cardiovascular Disease: Can We Trust Medico-administrative Hospital Data?

Laurie Lambert PhD , Claudia Blais PhD ^{b c}, Denis Hamel MSc ^b, Kevin Brown MSc ^a, Stéphane Rinfret MD, SM ^d, Raymond Cartier MD ^e, Maude Giguère ^a, Céline Carroll BSc ^a, Christiane Beauchamp ^a, Peter Bogaty MD ^d

I – Les BD du Québec

Banques de données créées à l'origine pour le remboursement des actes médicaux

1. MED-ÉCHO: séjour hospitalier, diagnostics, services...
2. FIPA: âge, régime public d'assurance médicaments...
3. Services pharmaceutiques: medicament, dose, durée...
4. Services médicaux: acte médical, diagnostics...
5. RED: date, cause principale...

Services pharmaceutiques
Médicaments

Services médicaux
Médecins

MED-ÉCHO
Hospitalisations

FIPA
Assurance médicaments

RED
Décès

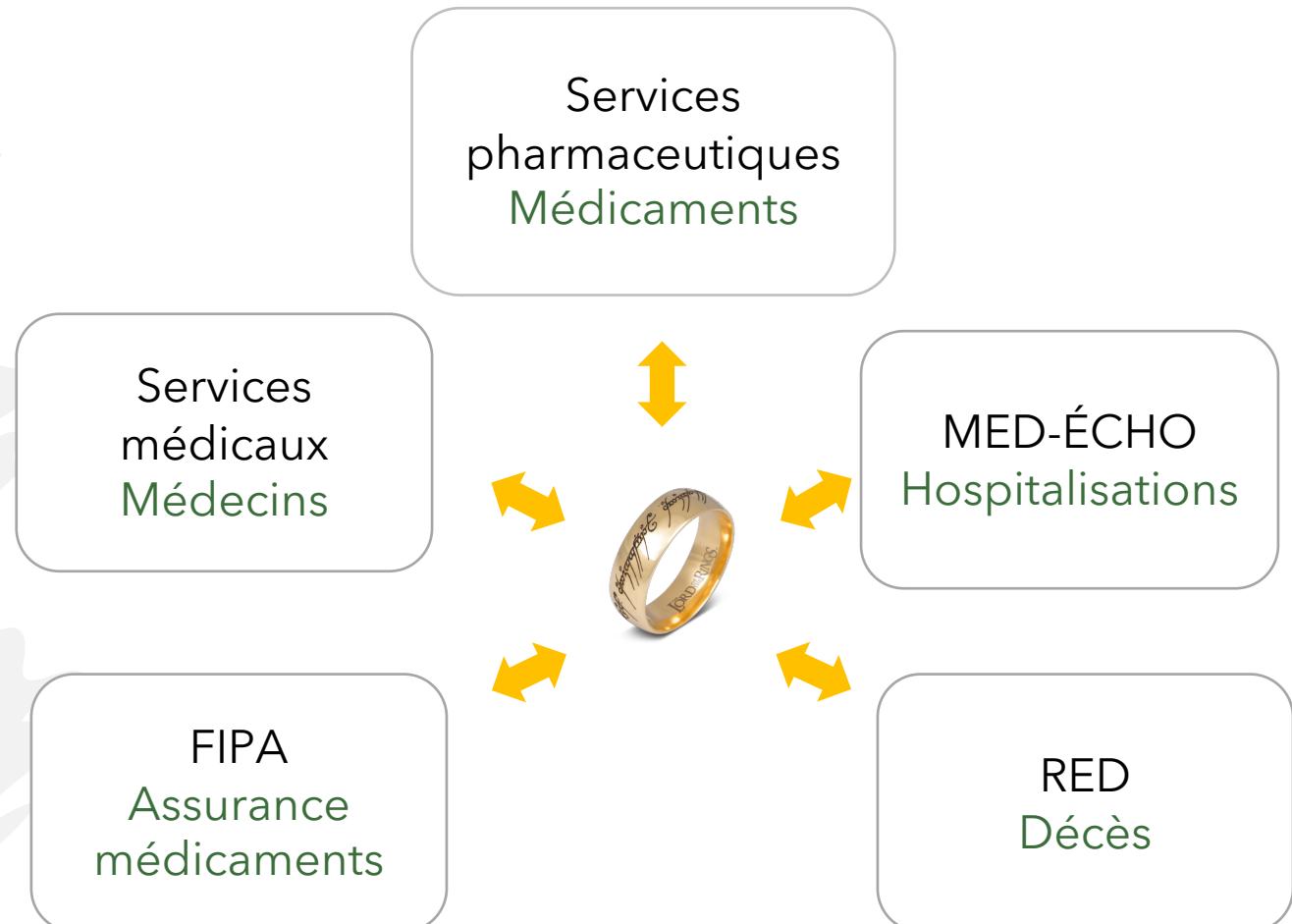
I – SISMACQ

Jumelage par un identifiant unique

→ Système intégré de surveillance des maladies chroniques du Québec

Hébergé dans un environnement sécurisé à l'INSPQ

Études observationnelles ou surveillance populationnelle de maladies chroniques



- I. Les banques de données médico-administratives pour la recherche
 - Avantages et inconvénients
 - SISMACQ
- II. Grande utilisation des services d'urgence
 - Profils des grands utilisateurs
 - Construction d'un outil de dépistage
- III. Surveillance de la polypharmacie
 - Fouille de données
 - Prédiction
 - Transparence
- IV. Conclusions et réflexions
 - On n'a pas parlé de...
 - Statistiques et IA pour la recherche en santé

II – Grande utilisation des SU

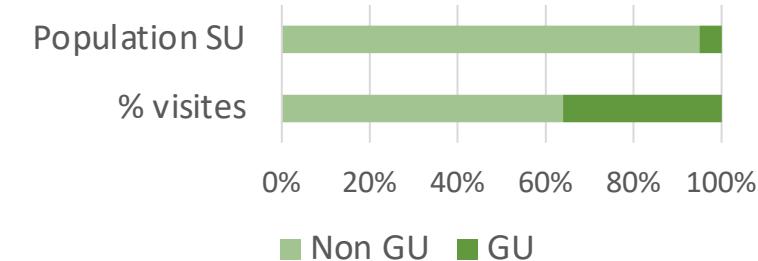
≥4 visites à l'urgence par année¹

- ➔ soins fragmentés, épisodiques
- ➔ qualité sous-optimale

Gestion de cas jugée efficace pour prévenir la grande utilisation²

Prédiction de la grande utilisation ~3%

- ➔ « débordement »³




Journal of Medical Engineering & Technology

Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: <https://www.tandfonline.com/loi/jmet20>

Machine learning in primary care: potential to improve public health

Jungwoo Kang , Moghees Hanif , Eushaa Mirza , Muhammad Asad Khan & Muzaffar Malik

nature > scientific reports > articles > article

Article | [Open access](#) | Published: 03 February 2023

Machine learning to improve frequent emergency department use prediction: a retrospective cohort study

[Yohann M. Chiu](#)✉, [Josiane Courteau](#), [Isabelle Dufour](#), [Alain Vanasse](#) & [Catherine Hudon](#)

1. Krieg, C., et al. (2016). Individual predictors of frequent emergency department use: a scoping review. *BMC health services research*, 16, 1-10.
2. Hudon, C., et al. (2016). Effectiveness of case management interventions for frequent users of healthcare services: a scoping review. *BMJ open*, 6(9), e012353.
3. Chiu, Y. M., et al. (2023). Machine learning to improve frequent emergency department use prediction: a retrospective cohort study. *Scientific Reports*, 13(1), 1981.

II – Grande utilisation des SU

Nécessité de bien comprendre les besoins car profils hétérogènes

Méthodes d'apprentissage non supervisé pourrait nous aider à définir des profils cliniquement significatifs?

Méthodes d'apprentissage supervisé pourrait nous aider à définir les facteurs prédictifs de la grande utilisation?

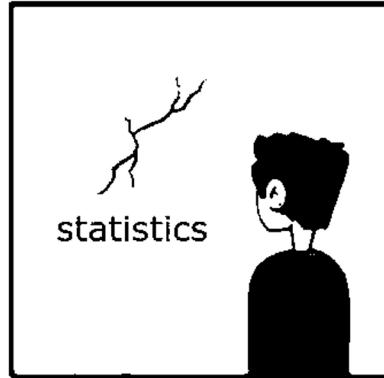
Variable		Total (%)	Non grands utilisateurs (%)	Grands utilisateurs (%)
Total		451,775 (100)	438,099 (100)	13,676 (100)
Femme		234,320 (51.9)	226,968 (51.8)	7,352 (53.8)
Âge	18-34	23,723 (5.3)	22,775 (5.2)	948 (6.9)
	35-54	83,393 (18.5)	80,977 (18.5)	2,416 (17.7)
	55-64	99,136 (21.9)	96,618 (22.1)	2,518 (18.4)
	65-74	116,323 (25.7)	113,198 (25.8)	3,125 (22.9)
	75-84	93,091 (20.6)	89,887 (20.5)	3,204 (23.4)
	≥ 85	36,109 (8.0)	34,644 (7.9)	1,465 (10.7)
Charlson	0	277,798 (61.5)	272,919 (62.3)	4,879 (35.7)
	1-2	98,228 (21.7)	94,558 (21.6)	3,670 (26.8)
	3-4	34,395 (7.6)	32,248 (7.4)	2,147 (15.7)
	≥ 5	41,354 (9.2)	38,374 (8.8)	2,980 (21.8)
Douleurs chroniques		75,263 (16.7)	71,859 (16.4)	3,404 (24.9)
Troubles mentaux		15,778 (3.5)	14,281 (3.3)	1,497 (10.9)
Hospitalisation		191,862 (42.5)	182,364 (41.6)	9,498 (69.5)

II – Grande utilisation des SU

Vocabulaire: apprentissage automatique (*machine learning*)

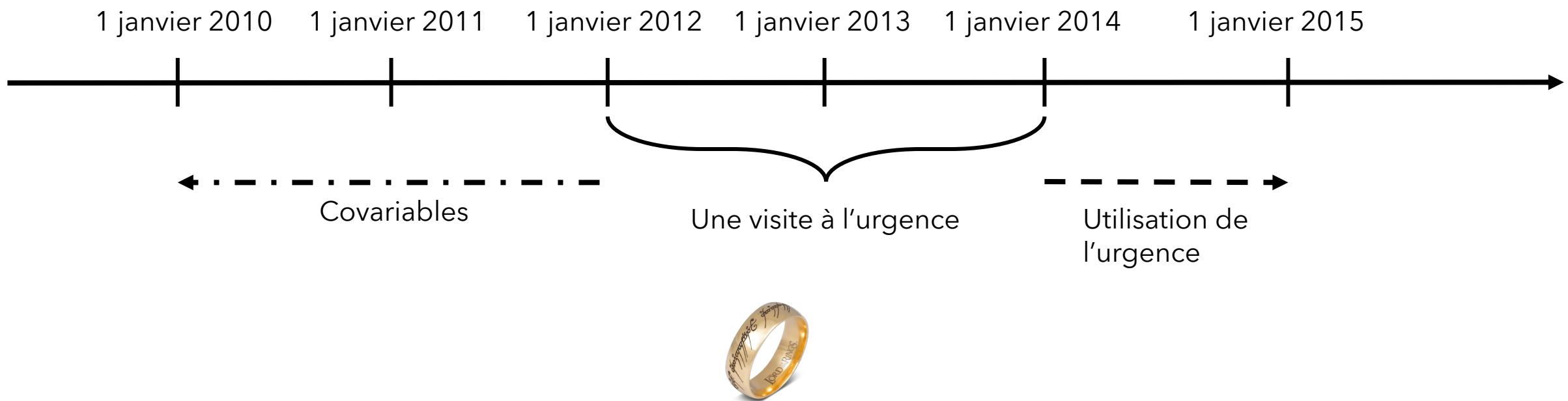
- Apprentissage supervisé (prédiction d'issues de santé)
- Apprentissage non supervisé (profils d'individus)
- Apprentissage par renforcement (un chat est un chat)

Beaucoup de concepts se recoupent entre les biostatistiques et l'apprentissage automatique



Adapté de <https://towardsdatascience.com/no-machine-learning-is-not-just-glorified-statistics-26d3952234e3>

II – Grande utilisation des SU



II – Grande utilisation des SU

Analyse de classes latentes: modèle probabiliste qui identifie des sous-groupes de population

Les individus au sein d'un même groupe sont similaires, tout en étant différents à ceux des autres groupes

1) Faible morbidité , 2) Comorbidité élevée, 3) Santé mentale ou abus de substances, 4) Blessures ou douleurs

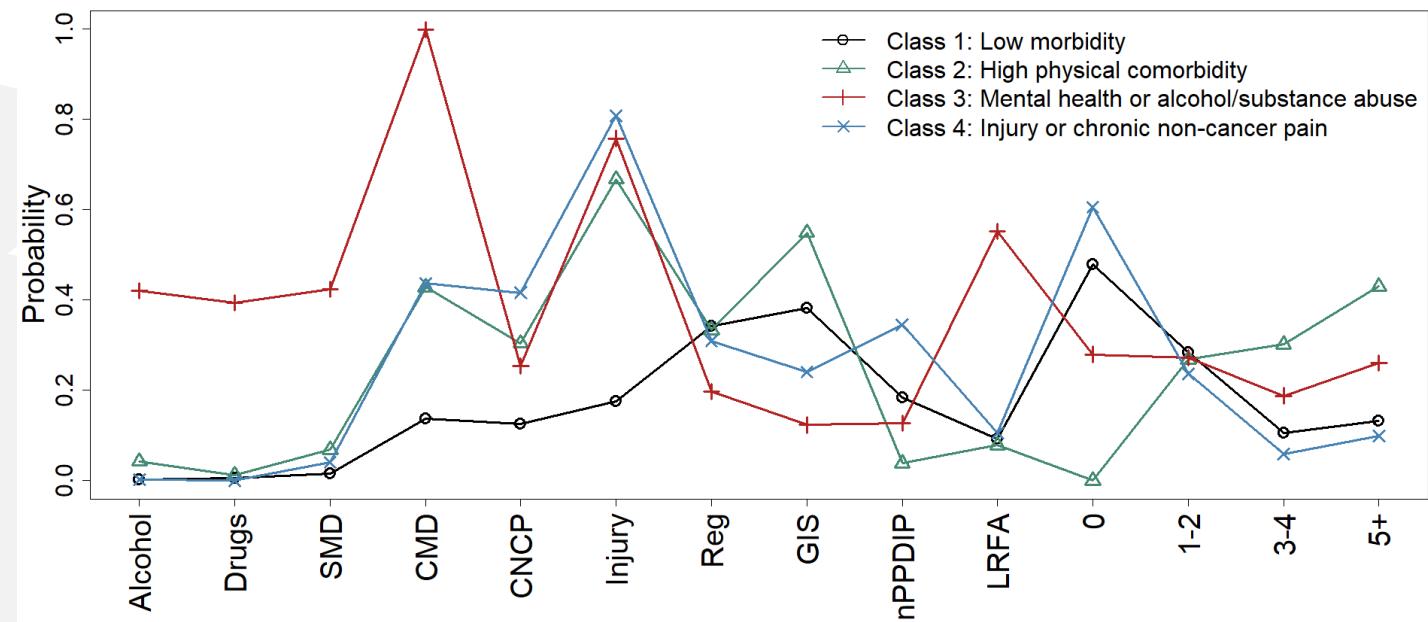
1. Chiu, Y. M., et al. (2022). Profiles of frequent emergency department users with chronic conditions: a latent class analysis. *BMJ open*, 12(9), e055297.

Open access

Original research

BMJ Open Profiles of frequent emergency department users with chronic conditions: a latent class analysis

Yohann Moanahere Chiu ^{1,2}, Isabelle Dufour,^{3,4} Josiane Courteau,² Alain Vanasse,^{1,2} Maud-Christine Chouinard,⁵ Marie-France Dubois,⁶ Nicole Dubuc,^{3,7} Nicolas Elazhary,¹ Catherine Hudon ^{1,2}



SMD: Serious mental disorders, CMD: Common mental disorders, CNCP: Chronic non-cancer pain. Reg, GIS, nPPDIP, and LRFA refer to PPDIP status while 0, 1-2, 3-4, and 5+ refer to the comorbidity index.

II – Grande utilisation des SU

Régression logistique pour identifier les facteurs cliniques les plus importants

Jumelage avec des questionnaires auto-rapportés (facteurs culturels et environnementaux)

Intégration des résultats dans un **outil de dépistage** des patients aux besoins complexes

1. Hudon, C., Bisson, M., Dubois, M. F., Chiu, Y., Chouinard, M. C., Dubuc, N., ... & Vanasse, A. (2021). CONECT-6: a case-finding tool to identify patients with complex health needs. *BMC health services research*, 21, 1-9.

[Home](#) > [BMC Health Services Research](#) > Article

CONECT-6: a case-finding tool to identify patients with complex health needs

Research article | [Open access](#) | Published: 17 February 2021

Volume 21, article number 157, (2021) [Cite this article](#)

Questions	Answers	
1. In general, would you say your health is fair or even poor?	Yes	No
2. Do you have pain or discomfort preventing most of your activities?	Yes	No
3. In the past 12 months, do you consider your health needs were met less than half of the time?	Yes	No
4. Do your interactions with the health system and health professionals ever make you feel like you have complex health problems?	Yes	No
5. Is your household income from all sources before taxes and other deductions less than \$20,000?	Yes	No
6. In the past 12 months, have you rarely or even never received support from friends or relatives when you needed it?	Yes	No
Number of yes and no answers	—	—

- I. Les banques de données médico-administratives pour la recherche
 - Avantages et inconvénients
 - SISMACQ
- II. Grande utilisation des services d'urgence
 - Profils des grands utilisateurs
 - Construction d'un outil de dépistage
- III. Surveillance de la polypharmacie
 - Fouille de données
 - Prédiction
 - Transparence
- IV. Conclusions et réflexions
 - On n'a pas parlé de...
 - Statistiques et IA pour la recherche en santé

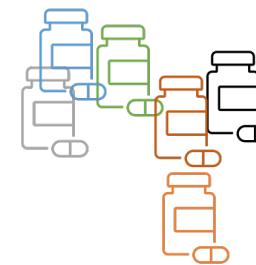
III – Surveillance de la polypharmacie

Polypharmacie associée à des **issues de santé négatives** (décès, fragilité, CHSLD, hospitalisation)¹

Surveillance accrue chez les populations vulnérables²

Difficile à départager étant donné le nombre impressionnant de combinaisons possibles

→ **méthodes IA?**



3500 dénominations communes



{Rx1; Rx2}



$6 \cdot 10^6$ possibilités

{Rx1; ...; Rx3}



$7 \cdot 10^9$ possibilités

{Rx1; ...; Rx4}



$6 \cdot 10^{12}$ possibilités

{Rx1; ...; Rx5}



$4 \cdot 10^{15}$ possibilités

1 combinaison/seconde ~ 130 millions d'années



1. Pazan, F., & Wehling, M. (2021). Polypharmacy in older adults: a narrative review of definitions, epidemiology and consequences. *European geriatric medicine*, 12, 443-452.
2. Gosselin, M., Talbot, D., Simard, M., Chiu, Y. M., ... & Sirois, C. (2023). Classifying Polypharmacy According to Pharmacotherapeutic and Clinical Risks in Older Adults: A Latent Class Analysis in Quebec, Canada. *Drugs & Aging*, 40(6), 573-583.

III – Surveillance de la polypharmacie

L'IA promet des résultats plus précis et plus rapides en épidémiologie

TRÈS populaire en médecine personnalisée... et en surveillance?

Deux méthodes de fouille de données pour déterminer les **séquences** et les **combinaisons fréquentes**

1. Bukhtiyarova, O., Abderrazak, A., Chiu, Y., Sparano, S., Simard, M., & Sirois, C. (2022). Major areas of interest of artificial intelligence research applied to health care administrative data: a scoping review. *Frontiers in Pharmacology*, 13, 944516.

Georg Thieme Verlag KG Stuttgart

Artificial Intelligence in Public Health and Epidemiology

Rodolphe Thiébaut, Frantz Thiessard, Section Editors for the IMIA Yearbook Section on Public Health and Epidemiology Informatics

[Author Affiliations](#)

Challenges and opportunities for public health made possible by advances in natural language processing

Oliver Bacic,^{1,*} Matthew Tunis,¹ Kelsey Young,¹ Coraline Doan,² Howard Swerdfeger,² and Justin Schonfeld^{3,*}

[Author information](#) [Copyright and License information](#) [Disclaimer](#)

¹ Faculty of Pharmacy, Université Laval, Québec, QC, Canada

² Quebec Nation

³ Faculty of Medi

Home » American Journal of Public Health (AJPH) » January 2021

Artificial Intelligence, Intersectionality, and the Future of Public Health

Greta R. Bauer PhD, MPH, and Daniel J. Lizotte PhD

[+] Author affiliations, information, and correspondence details

Accepted: October 09, 2020 Published Online: December 16, 2020

Review Article | Open Access | Published: 26 February 2021

Applications of digital health for public health responses to COVID-19: a systematic scoping review of artificial intelligence, telehealth and related technologies

Dinesh Visva Gunasekeran, Rachel Marjorie Wei Wen Tseng, Yih-Chung Tham & Tien Yin Wong

npj Digital Medicine 4, Article number: 40 (2021) | [Cite this article](#)

11k Accesses | 48 Citations | 47 Altmetric | [Metrics](#)

Conclusions: The scoping review revealed the potential of AI application to health-related studies. However,

several areas of interest in pharmacoepidemiology are sparsely reported, and the lack of details in studies related to pharmacotherapy suggests that AI could be used more optimally in pharmacoepidemiologic research. Finally, the performance of digital health technology for operational applications related to **population surveillance** [...] have not been adequately evaluated"

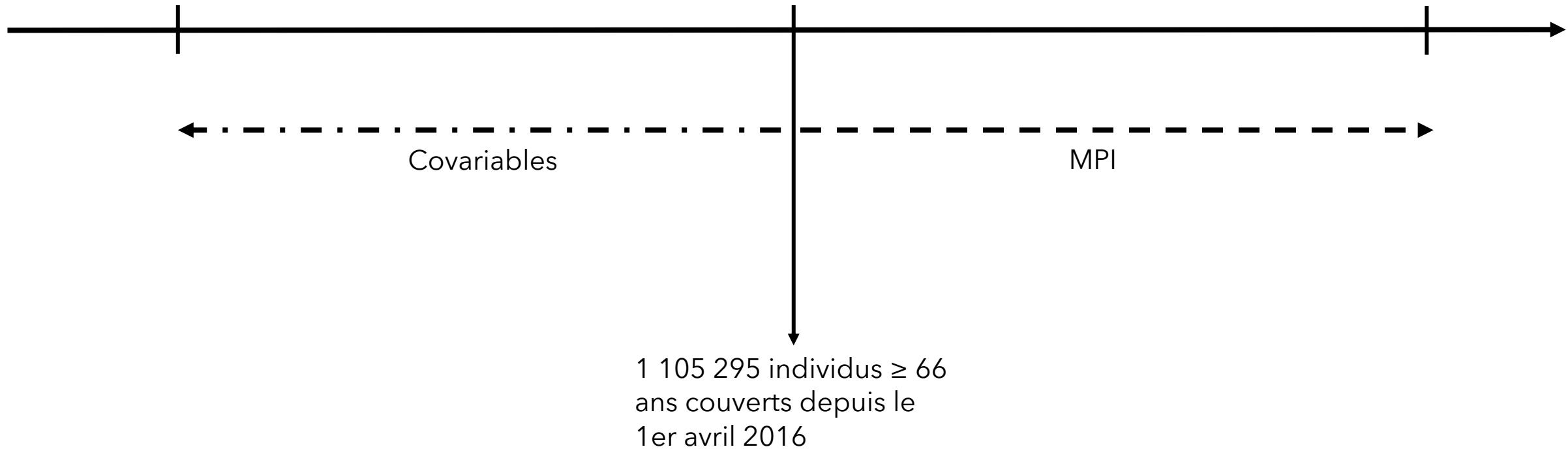
III – Surveillance de la polypharmacie



1 avril 2014

1 avril 2015

1 avril 2016

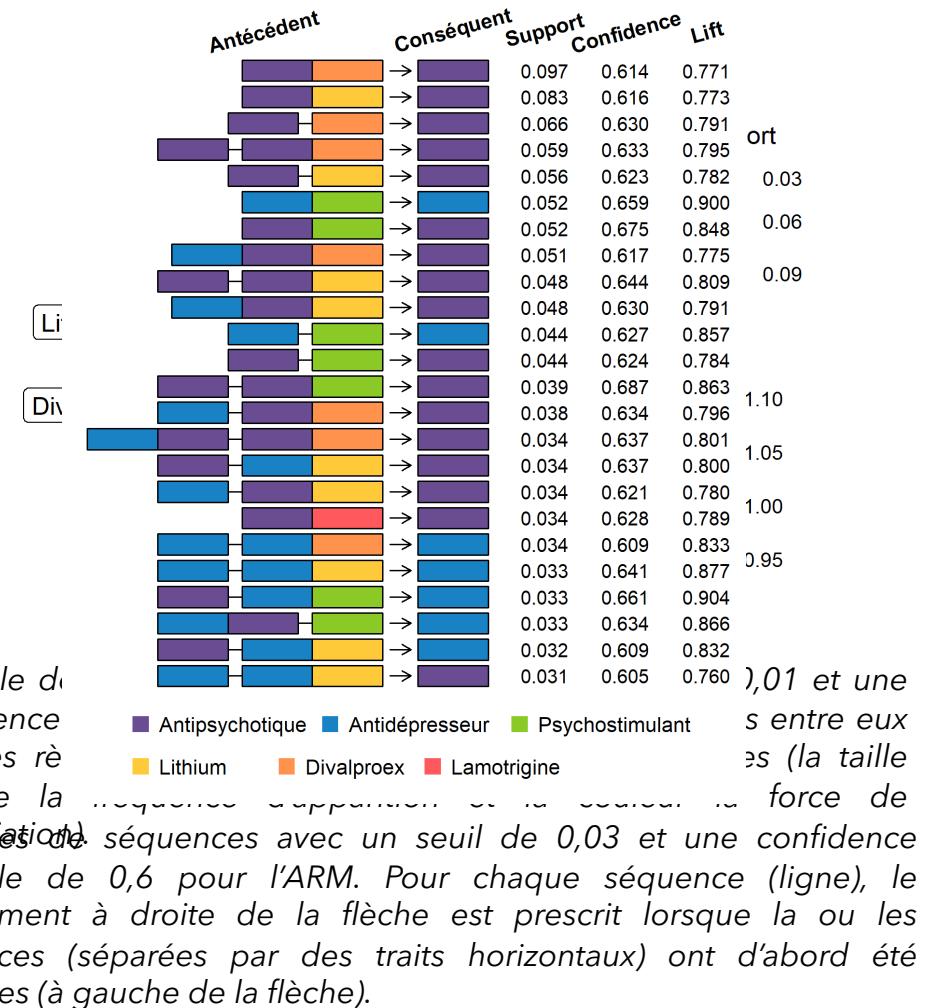


III – Polypharmacie et fouille de données

Combinaison ou séquence fréquente se mesure avec le **support**

Prescrire du lithium augmente de 10% la probabilité de prescrire un antipsychotique (**lift**)

Antipsychotique est prescrit dans 61% des cas lorsque la combinaison antipsychotique + divalproex est d'abord prescrite (**confidence**)



III – Polypharmacie et prédiction

Régression logistique (RL): modèle de régression pour variable binaire

Gradient boosting machines (GBM): modèles qui utilisent ensemble de « couches » de médicaments potentiellement inappropriés (NPI) : application du théorème de Bayes

Réseaux de neurones (RN): ajustes pour l'âge, le sexe, la situation de résidence, les maladies chroniques, le nombre

Forêts aléatoires (FA): arbres de décisions successifs

Méasures sous la courbe ROC, sensibilité, spécificité, valeurs prédictives positives et négatives

Modèle	AUC 95% CI	Sens	Spéc	VPP	VPN	Préc	Brier
RL	0.62 0.62 - 0.62	0.49	0.73	0.62	0.61	0.61	0.38
GBM	0.62 0.62 - 0.62	0.49	0.73	0.63	0.61	0.61	0.38
NB	0.61 0.61 - 0.61	0.38	0.80	0.63	0.59	0.59	0.40
RN	0.62 0.62 - 0.62	0.52	0.70	0.62	0.62	0.61	0.38
FA	0.62 0.62 - 0.62	0.51	0.72	0.62	0.62	0.61	0.38

III – Polypharmacie et prédiction

Qu'en est-il de l'**importance** des variables dans la prédiction?

Différentes mesures pour l'importance selon le modèle

Importance **globale** semble concordante~ish

- Chiu, YM., Sirois, C., Simard, M., Gagnon, ME. & Talbot, D. (2024). Traditional methods hold their ground against machine learning in predicting potentially inappropriate medication use in older adults. *Value in Health*, revisions required.

Variable	RL	FA	GBM	NB	RN
Troubles mentaux	1	1	1	7	15
Sexe	2	2	2	2	2
Diabète	3	4	5	6	12
Hypertension	4	5	3	1	7
MPOC	5	3	4	4	10
Hospitalisation	6	7	6	9	6
MCV	7	9	7	5	8
Alzheimer	8	6	10	14	14
Asthme	9	8	9	10	11
Zone résidentielle	10	10	12	15	3
Âge	11	11	8	3	1
Ostéoporose	12	12	11	8	13
Défavorisation sociale	13	14	14	11	5
Insuffisance cardiaque	14	13	13	12	9
Défavorisation matérielle	15	15	15	13	4

+vert = +important

III – Polypharmacie et prédiction

Pas de vainqueur clair pour les performances...

On n'a pas pris en compte les facteurs liés à l'utilisation des médicaments et le style de vie

Dans la boite noire... d'autres boites noires

Comment expliquer la performance des RN?



Variable	RL	FA	GBM	NB	RN
Troubles mentaux	1	1	1	7	15

III – Polypharmacie et explicabilité

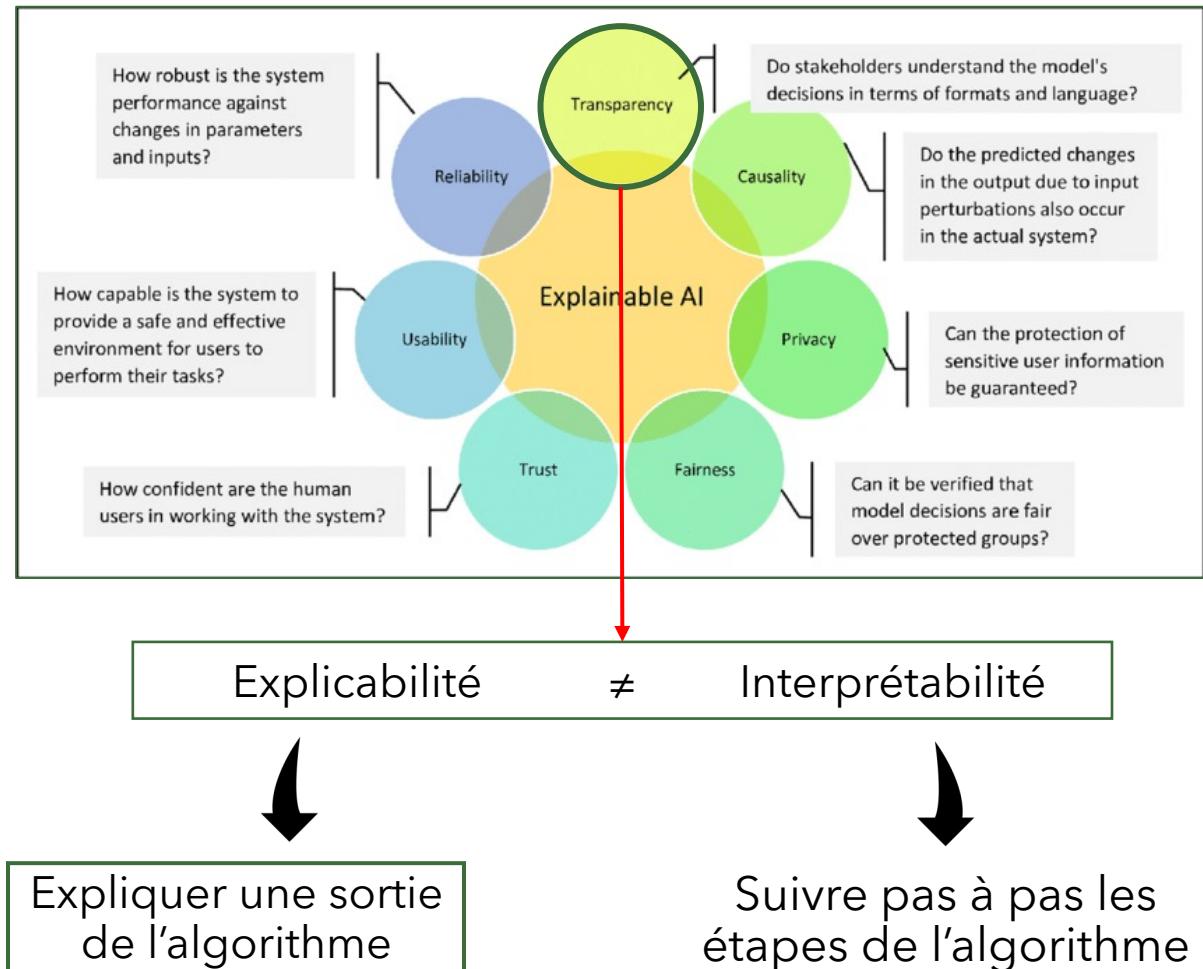
IA ≈ boîte noire

Transparence: deux caractéristiques différentes en XAI

Être capable d'expliquer une prédiction de l'algorithme aide à la **généralisabilité/confiance**

Rendre la boîte noire... plus grise?

Source: Fiok, K., Farahani, F. V., Karwowski, W., & Ahram, T. (2022). Explainable artificial intelligence for education and training. *The Journal of Defense Modeling and Simulation*, 19(2), 133-144.



III – Polypharmacie et explicabilité

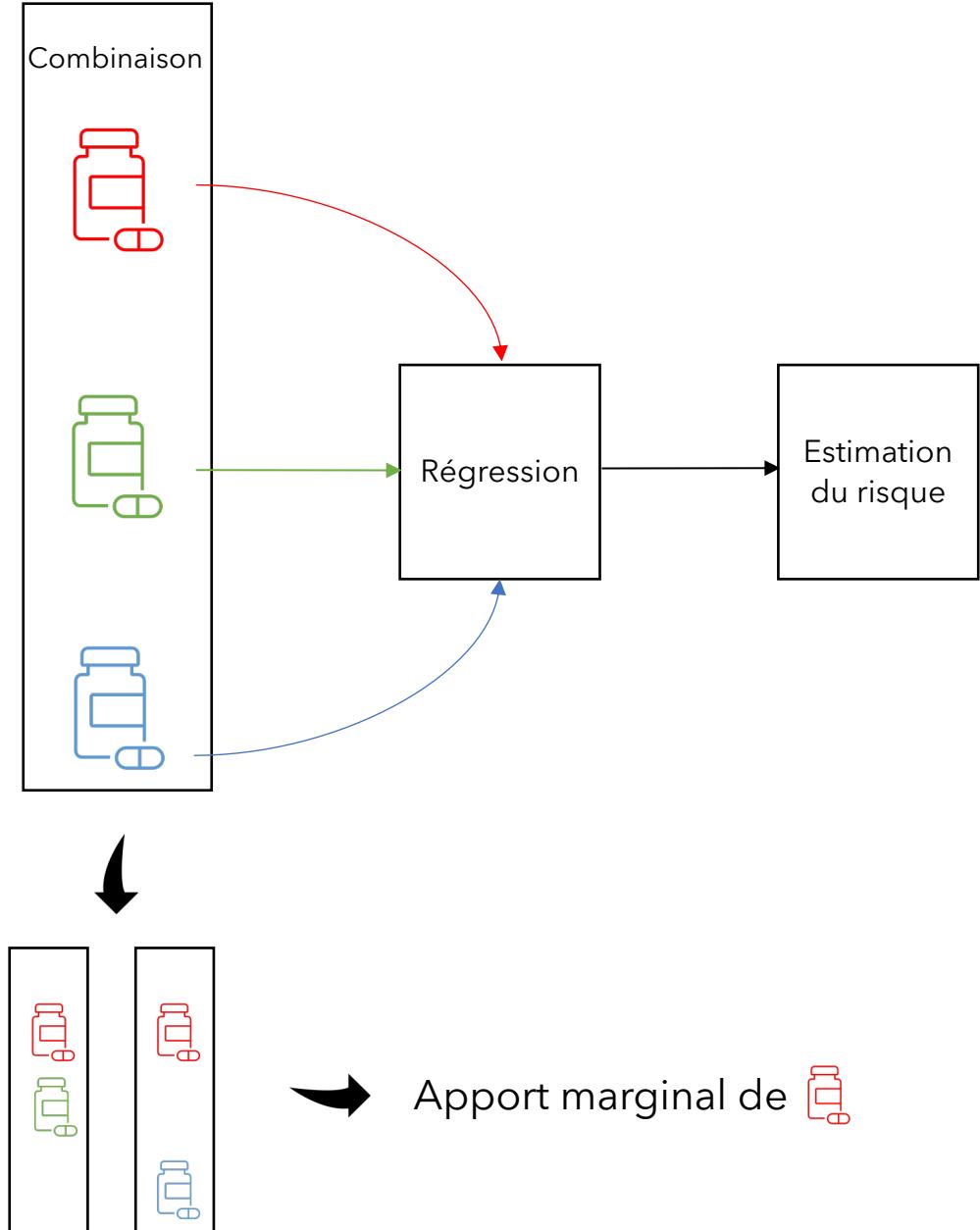
Valeurs SHAP¹

Théorie des jeux: dans une équipe de N joueurs possible, quel est l'apport du joueur i dans le gain total ν

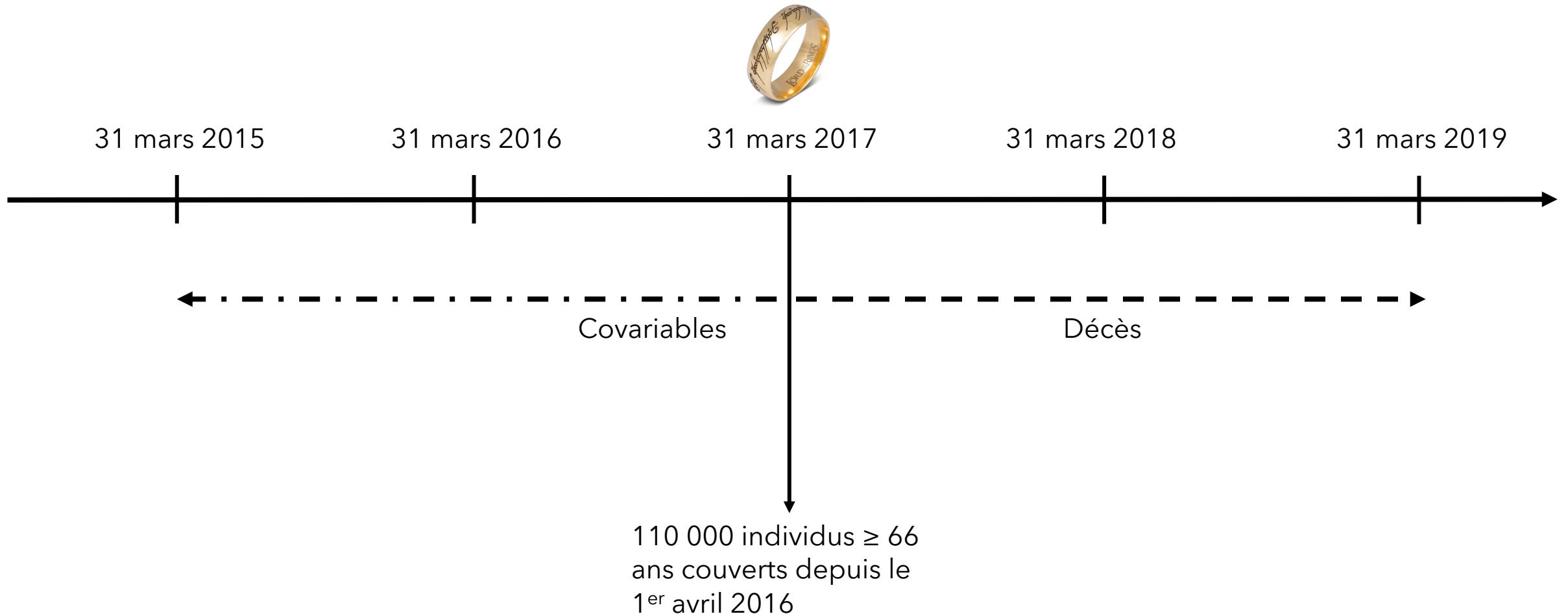
Il faut prendre en compte:

- Le « poids » de chaque joueur
- Les interactions entre les joueurs
- L'ordre d'arrivée des joueurs

1. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 30.



III – Polypharmacie et explicabilité

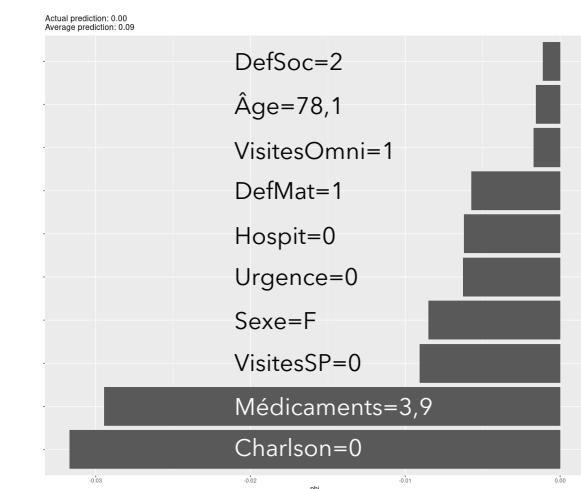
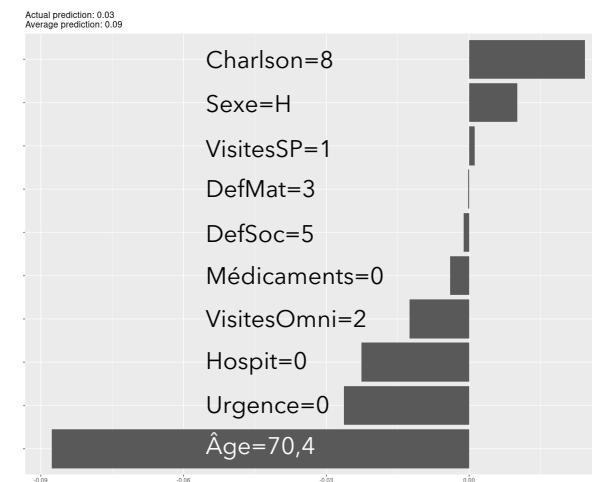
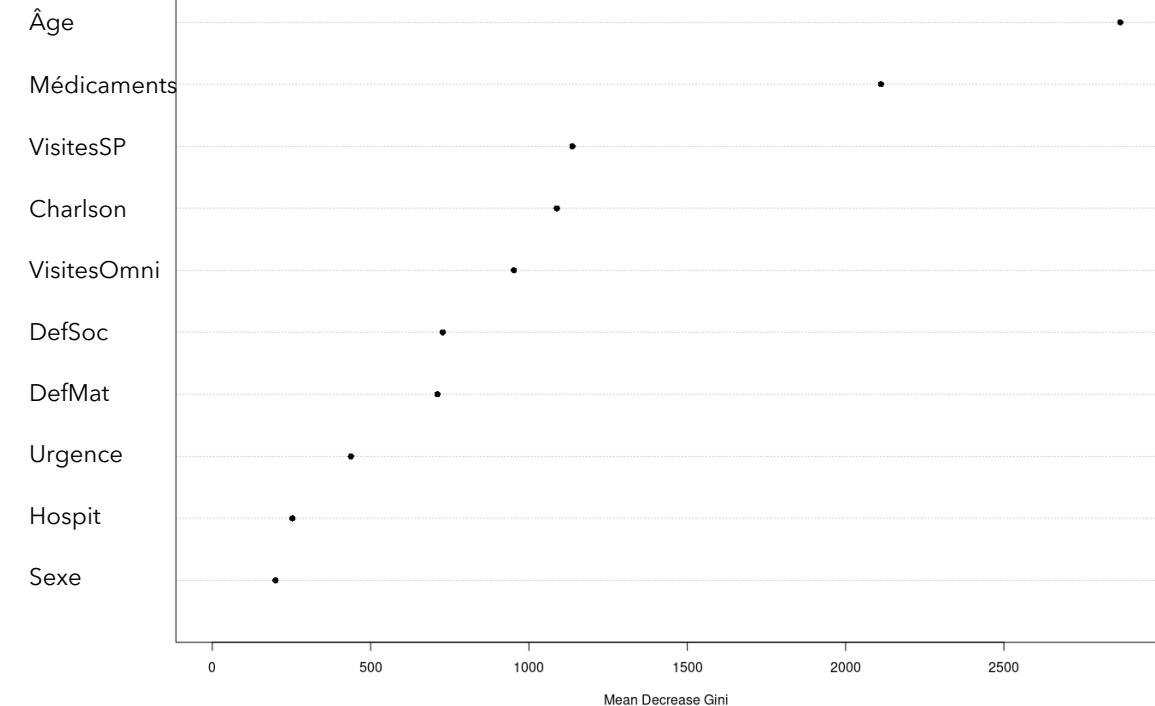


III – Polypharmacie et explicabilité

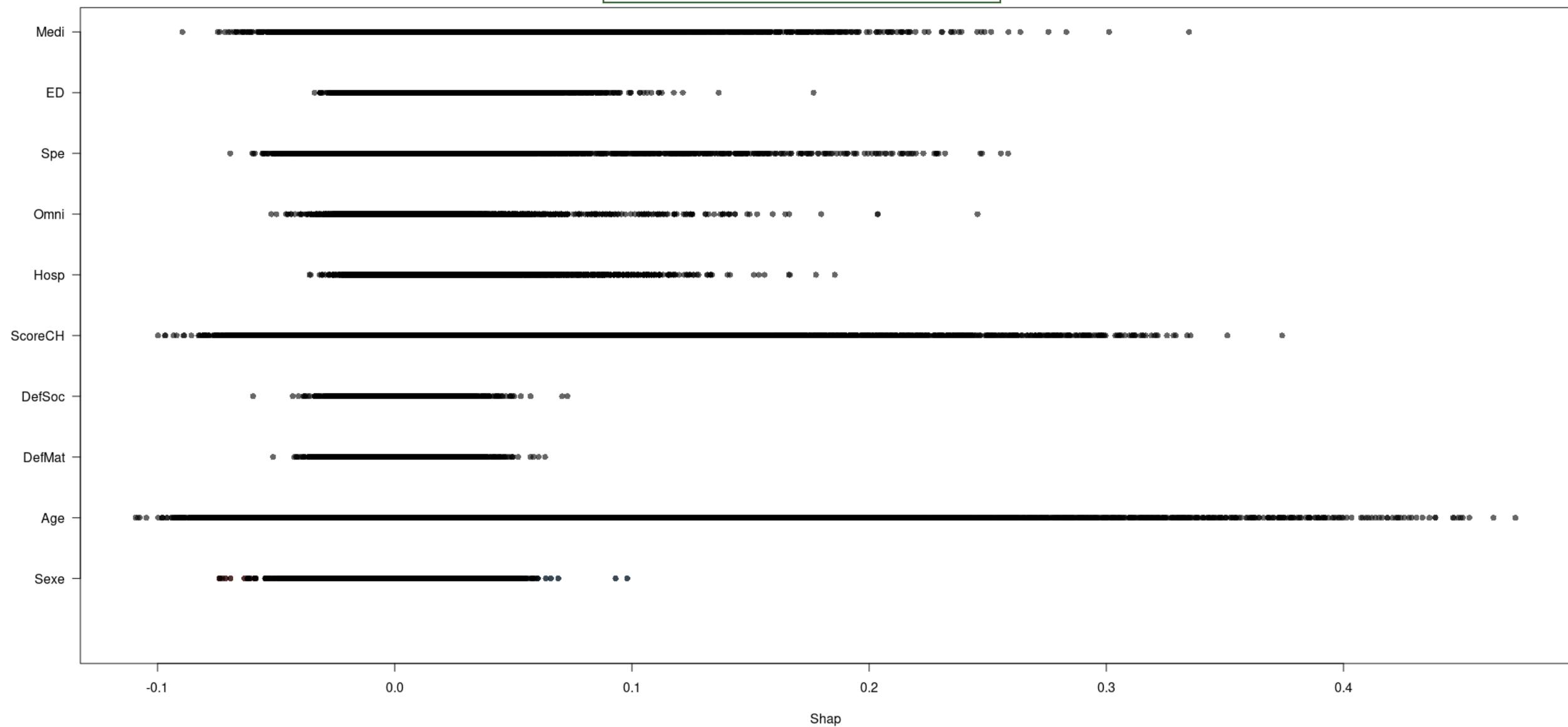
Tous les modèles AI/ML/stats mesurent l'importance **globale**

SHAP permet une inspection **locale** de l'importance, c'est-à-dire pour un individu

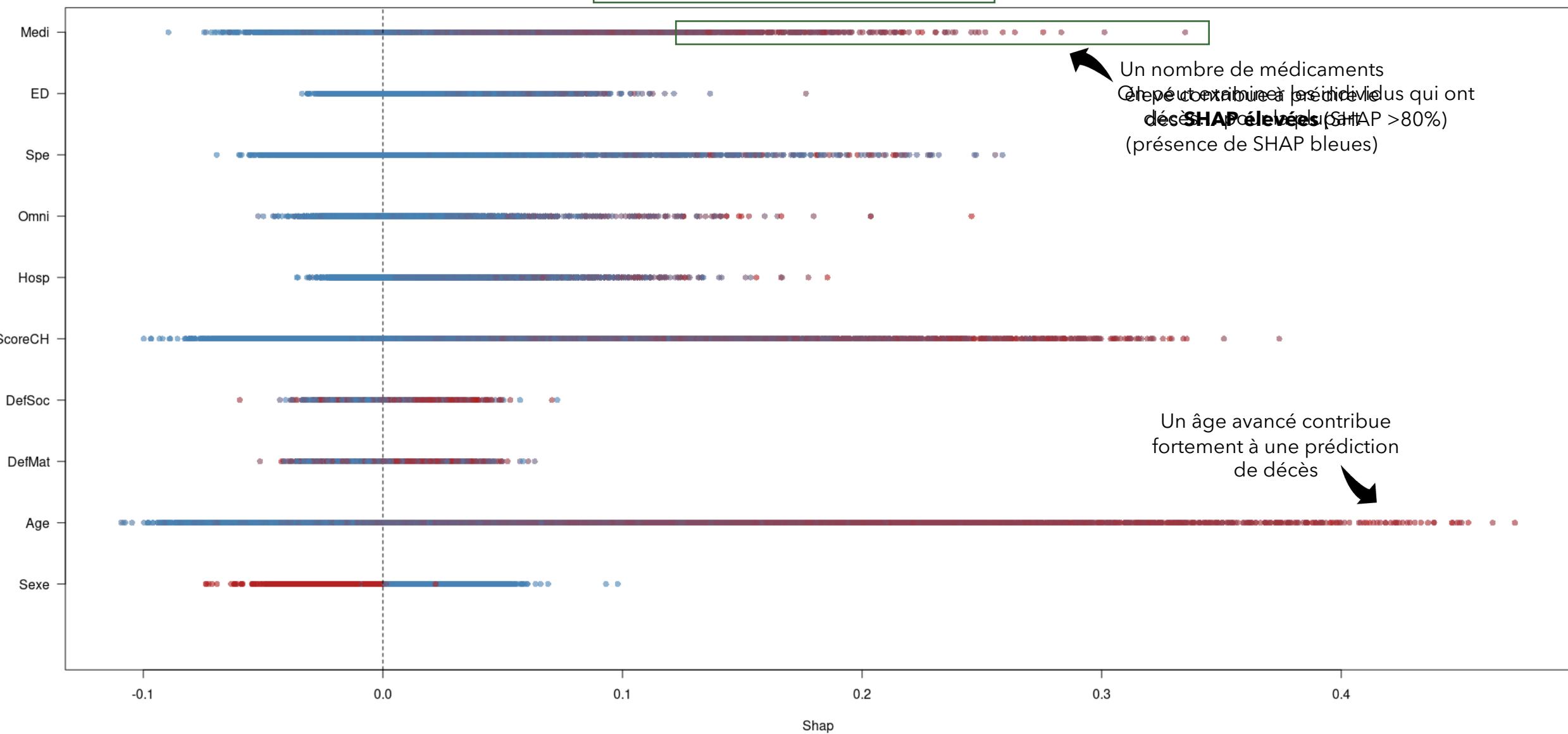
Quelles valeurs de quelles variables ont eu un impact sur la prédiction?



Valeur SHAP élevée « pousse » le
modèle à prédire le décès



Rouge: valeurs élevées de la variable
Bleu: valeurs basses de la variable



III – Polypharmacie et explicabilité

La population SHAP>80% a des **caractéristiques similaires** avec une population >8 médicaments

SHAP>80% est plus intéressante puisqu'il y a des individus qui ont 0, 1, 2... médicaments

SHAP permet également d'observer la forme de la relation du pouvoir prédictif

Variable (moyenne)	Population générale	SHAP > 80%
Age	75,6	78,6
Femme %	55,6	55,5
Score de Charlson	1,6	3,3
Nombre d'hospitalisations	0,1	0,3
Nombre de visites à l'urgence	0,5	0,8
Nombre de visites médecin spécialiste	4,3	6,7
Nombre de médicaments - médiane (Q1-Q3)	4,6 (2,1-7,3)	10,1 (8,5-12,3)

III – Polypharmacie et explicabilité

SHAP

- Permet d'examiner l'hétérogénéité dans la prédiction
- Long... très long: 2 jours de calcul pour ~30'000 individus

Il existe d'autres méthodes d'explicabilité des modèles IA

Chacune présente des défis et avantages distincts

Aucune n'est une fin en soi, mais toute méthode d'explicabilité est utile pour la prédiction

scientific reports

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [scientific reports](#) > [articles](#) > [article](#)

Article | [Open access](#) | Published: 04 April 2023

Explanatory predictive model for COVID-19 severity risk employing machine learning, shapley addition, and LIME

[Mariam Laatifi](#), [Samira Douzi](#)✉, [Hind Ezzine](#), [Chadia El Asry](#), [Abdellah Naya](#), [Abdelaziz Bouklouze](#), [Younes Zaid](#) & [Mariam Naciri](#)

[Scientific Reports](#) 13, Article number: 5481 (2023) | [Cite this article](#)

1875 Accesses | 4 Citations | 5 Altmetric | [Metrics](#)

- I. Les banques de données médico-administratives pour la recherche
 - Avantages et inconvénients
 - SISMACQ
- II. Grande utilisation des services d'urgence
 - Profils des grands utilisateurs
 - Construction d'un outil de dépistage
- III. Surveillance de la polypharmacie
 - Fouille de données
 - Prédiction
 - Transparence
- IV. Conclusions et réflexions
 - On n'a pas parlé de...
 - Statistiques et IA pour la recherche en santé

IV – Conclusions et réflexions



On n'a pas parlé...

- D'apprentissage fédéré
- Des enjeux de protection de la vie privée, d'éthique
- Des limites liées à la manipulation des banques de données

En résumé...

- Les banques de données médico-administratives offrent des opportunités uniques pour la recherche populationnelle
- Il existe beaucoup d'outils d'analyse statistique (BEAUCOUP)
- L'IA gagne en popularité



DEUXIÈME SESSION

QUARANTE-DEUXIÈME LÉGISLATURE

Projet de loi n° 19

Loi sur les renseignements de santé et de services sociaux et modifiant diverses dispositions législatives

Présentation

Présenté par
M. Christian Dubé
Ministre de la Santé et des Services sociaux

Éditeur officiel du Québec
2021

INSPQ INSTITUT NATIONAL DE SANTÉ PUBLIQUE DU QUÉBEC



Cadre de qualité des données du Système intégré de surveillance des maladies chroniques du Québec

RAPPORT MÉTHODOLOGIQUE



Québec

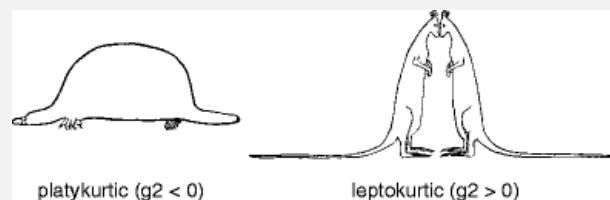
« L'ampleur du SISMACQ, qui contient plusieurs millions d'informations dans une **vingtaine de bases de données** structurées selon un modèle relationnel complexe, engendre plusieurs défis pour son utilisation. Le **temps d'exécution et l'espace mémoire** disponible représentent aussi des défis d'utilisation »

IV – Conclusions et réflexions

Les méthodes IA nécessitent de la réflexion en amont de l'application

Par ex. : Vocabulaire différent

Par ex. bis: En statistique, on suppose des modèles (omniprésence de la loi normale); en IA/ML, on laisse « parler les données »

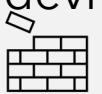


Wright, D.B., Herrington, J.A. Problematic standard errors and confidence intervals for skewness and kurtosis. *Behav Res* 43, 8-17 (2011).

Biostatistique Épidémiologie	AI/ML
Estimation	Apprentissage
Régression	Apprentissage supervisé
Sensibilité/Valeur prédictive positive	Rappel/Précision
Variable dépendante	Cible
Paramètres	Poids
Variable indépendante	Caractéristique (<i>feature</i>)
Observation	Instance
	Aire sous la courbe ROC
	Surajustement
	Modèle

IV – Conclusions et réflexions

L'application de IA/ML ne devrait pas être un but ultime



On augmente la boîte (noire) à outils



Pas de gain (majeur) de performances avec des données traitées et structurées

Surtout: besoin de s'accorder sur des objectifs communs!

Biostatistique Épidémiologie	AI/ML
Estimation	Apprentissage
Régression	Apprentissage supervisé
...	...
Expliquer associations	Estimer des paramètres
Santé des populations	Algorithmes optimaux
...	...
But commun!	

Merci! Avez-vous des questions pour moi?

Traductions libres

« Tous les modèles sont faux,
certains sont utiles » (George Box)

« Il n'est pas très utile de dire que tous les modèles sont faux [...] Ne pas tomber en amour d'un seul modèle pour exclure les autres » (Peter McCullagh et John Nelder)

« Les statistiques sont la grammaire des sciences » (Karl Pearson)