

Prédiction du nombre de patients hospitalisés atteints du SARS-CoV-2, apport des données publiques et de l'entrepôt de données de santé du CHU de Bordeaux.

Thomas Ferté^{a,b,c}, Vianney Jouhet^{c,d}, Romain Griffier^{c,d}, Boris P Hejblum^{a,b,e}, Rodolphe Thiébaud^{a,b,c,e}

a : Univ. Bordeaux, Centre Inserm Bordeaux Population Health, équipe SISTM, UMR 1219, 146 rue Léo Saignat, Bordeaux France

b : INRIA Bordeaux Sud Ouest, équipe SISTM, Talence, France

c : CHU Pellegrin, Groupe Hospitalier Pellegrin, Place Amélie Raba Léon, Bordeaux, France

d : Univ. Bordeaux, Inserm, Centre Inserm Bordeaux Population Health, équipe ERIAS, UMR 1219, F-33000 Bordeaux, France

e : Vaccine Research Institute, VRI, Hôpital Henri Mondor, 51 avenue du Maréchal de Lattre de Tassigny, Créteil, France

Introduction : Plusieurs sources de données relatives au suivi de l'épidémie de COVID-19 en France sont publiquement disponibles (e.g. *Reverse Transcription-Polymerase Chain Reaction* (RT-PCR), hospitalisations, variants dominant, données météorologiques, ...), et peuvent être utilisées pour prédire l'évolution de la charge hospitalière à court terme. Par ailleurs certains hôpitaux, comme celui du Centre Hospitalo-Universitaire (CHU) de Bordeaux, sont dotés d'un entrepôt de données de santé (EDS) qui permet d'obtenir des données agrégées supplémentaires (e.g. passages aux urgences, appels au Service d'Aide Médicale Urgente (SAMU)). Nous évaluons ici l'apport des données publiques et de l'EDS pour la prédiction à 7 et 14 jours du nombre de patients hospitalisés prévalent atteints du SARS-CoV-2, au sein du CHU de Bordeaux entre le 1^{er} décembre 2020 et le 16 août 2021.

Méthode : L'apport des sources de données a été évalué par quatre modèles de régression linéaire avec pénalisation elastic-net incluant des ensembles de prédicteurs différents et dont la variable à expliquer est nombre de patients hospitalisés à 7 jours : (i) données d'hospitalisations et de RT-PCR, en considérant le CHU de Bordeaux comme un département additionnel parmi les autres départements français, l'unité d'observation étant le département-jour (ii) idem que (i) en ajoutant les données de variants, de vaccination et météorologiques, (iii) idem que (ii) mais en considérant seulement les données du CHU et de la Gironde, l'unité d'observation étant le jour, (iv) idem que (iii) en ajoutant les données propres à l'EDS du CHU de Bordeaux. Pour l'étape (iv), la prédiction a été comparée à celle obtenue par une forêt aléatoire et une régression de poisson. Les performances de prédiction ont été évaluées du 1^{er} décembre 2020 au 16 août 2021 avec un ré-apprentissage quotidien sur les données disponibles du 16 mai 2020 jusqu'à la date de prédiction. Les performances de prédictions ont été comparées par un test de Wilcoxon pour données appariées. Ces mêmes analyses ont également été réalisées avec un horizon de 14 jours.

Résultats : Nos résultats suggèrent que l'ajout de l'information portant sur les variants, la météorologie ou la vaccination ne permet pas d'améliorer les performances prédictives avec une erreur absolue médiane (EAM) entre les modèles (i) et (ii) passant respectivement de 9,40 à 9,08 hospitalisations à 7 jours ($p = 0.756$) et de 16,52 à 19,08 hospitalisations à 14 jours ($p = 0.057$). L'ajout des données de l'EDS permet une amélioration de l'EAM entre le modèle (iii) et (iv) passant respectivement de 9,42 à 6,87 hospitalisations à 7 jours ($p = 0.087$) et de 11,54 à 9,15 hospitalisations à 14 jours ($p = 0.014$). A l'étape (iv), les variables les plus fréquemment sélectionnées étaient respectivement celles relatives aux hospitalisations, aux RT-PCR, au jour de la semaine et à la météorologie à 7 jours et celles relatives aux RT-PCR, au SAMU, aux urgences, et à la météorologie à 14 jours.

Discussion : L'ajout des données de l'EDS améliore les performances de prédictions du nombre prévalent d'hospitalisations comparé aux seules données disponibles publiquement.