

Hospitalisation et mortalité durant la crise Covid : Analyse de causalité

Dhalil Bello¹, Ulrich Segodo², Miracle Vodoumbo³

¹Etudiant en Master 1 Economie et en Magistère 2 Ingénieur Economiste à Aix-Marseille School of Economics

27 février 2023

Résumé

La crise sanitaire du Covid a eu une forte incidence sur la mortalité partout dans le monde. Les hôpitaux ont malheureusement connu un nombre relativement important de décès. Sur la base de cette observation et en travaillant avec des données du gouvernement mexicain, nous voulons vérifier si durant la forte crise, recevoir les soins à l'hôpital à la suite d'un diagnostic chronique augmente les chances de mourir. En contrôlant un certain nombre de caractéristiques importantes de la population hospitalisé et à domicile, nous supposons que la différence de mortalité entre ces deux échantillons est à attribuer au fait de rester à l'hôpital. En utilisant des méthodes d'appariement, nous observons un effet positif et significatif du fait de rester à l'hôpital sur la probabilité de décès.

Superviseur :
Mathieu Lefebvre



1 Introduction

La sévère crise sanitaire liée à l'épidémie de covid-19 a été soudaine et imprévisible. Les premiers foyers de contagions ont été identifiés dans la région de Wuhan en Chine, avant une rapide expansion dans la plupart des pays d'Europe et notamment en France. Selon LeFigaro (2022), la France serait le pays d'Europe, hors micro-états, avec l'incidence la plus élevée, 3.733 cas pour 100.000 habitants pour un total de 129.022 décès depuis le début de la pandémie au printemps 2020. Les nombres d'hospitalisation et de décès à l'hôpital ont considérablement augmenté avec l'expansion du virus surtout durant la première vague. Une étude menée par l'ANSM indique que sur la période d'octobre à décembre 2020, 88.940 patients sans antécédents d'hospitalisation pour Covid-19 ont été hospitalisés en raison de ce virus, dont 16.894 en sont décédés à l'hôpital. Bien que ces chiffres ne permettent pas de tirer un quelconque effet de causalité, il est avancé qu'en raison de l'état chronique de la plupart des patients hospitalisés, l'espérance de vie était plus faible dans les hôpitaux par rapport aux domiciles. Cependant, il a été souligné dans plusieurs publications que le nombre de décès à domicile était relativement beaucoup plus faible et que certains patients étaient atteints d'un niveau chronique du virus à domicile en l'ignorant et en le traitant comme une simple grippe. La plupart des personnes infectées par le virus COVID-19 souffriraient d'une maladie respiratoire légère et se rétabliraient sans nécessiter de traitement particulier. Les personnes âgées et celles qui présentent des problèmes médicaux sous-jacents tels que les maladies cardiovasculaires, le diabète, les maladies respiratoires chroniques et le cancer sont plus susceptibles de développer une maladie grave.

Dans ce contexte de panique généralisée liée à la hausse excessive du nombre d'hospitalisation et de décès à l'hôpital, existerait-il un effet causal entre l'hospitalisation à la suite d'un diagnostic chronique de covid et l'espérance de vie? Faire ressortir l'existence d'un tel lien, permettra de mieux réfléchir sur le type de ressources dont un individu pourrait avoir besoin au moment où il est testé positif sans nécessairement être hospitalisé et d'organiser les ressources nécessaires pour sauver la vie de ce patient.

En utilisant des données de patients fournies par le gouvernement mexicain pendant la crise de covid, l'objectif de notre étude est d'essayer de ressortir un lien causal entre la propension à mourir suivant qu'on est hospitalisé ou à domicile à la suite d'un diagnostic chronique de covid. Nous utilisons l'appariement par score de propension afin de construire un échantillon dans lequel les individus traités (étant atteint d'un niveau chronique et hospitalisé) et non traités (étant atteint d'un niveau chronique et à domicile) ont des caractéristiques similaires en termes d'âge, de sexe, de caractéristiques socio-démographiques individuelles (fumeur, alcoolique, etc.) et d'état de santé. Nous supposons qu'après avoir contrôlé les déterminants de l'hospitalisation à la suite d'un diagnostic chronique de covid, la différence de mortalité entre les deux groupes est à attribuer à l'organisation du traitement au sein des hôpitaux ou à l'état psychologique des patients une fois hospitalisé par rapport à l'alternative de recevoir les soins à leurs domiciles. Cela nous permet de déterminer s'il existe un lien de causalité entre le fait d'être hospitalisé et la probabilité de décès

lorsqu'on est atteint d'un niveau chronique de covid.

Après la présentation des données et de l'échantillon sélectionné ainsi que des statistiques descriptives en section 2, nous présenterons en section 3 la méthode d'appariement par score de propension utilisée pour surmonter le biais de sélection ainsi que les différents modèles que nous ferons ressortir à travers cette méthode pour s'assurer de la faible sensibilité de notre indicateur aux outliers et donc de sa robustesse. La section 4 présente les résultats des principales régressions effectuées ainsi que les analyses additionnelles de robustesse. La section 5 conclut.

2 Données et Statistiques descriptives

L'ensemble de données a été obtenu sur le site kaggle et provient des bases de données du gouvernement mexicain. Cet ensemble de données contient un nombre énorme d'informations anonymisées relatives aux patients, y compris les conditions préalables. L'ensemble de données brutes se compose de 21 caractéristiques uniques tels que : sexe, âge, classifications des tests covid, type de patient, pneumonie, grossesse, diabète, MPOC, asthme, immunodéprimé, hypertension, cardiovasculaire, autre maladie, obésité, rénale chronique, tabac, date de décès et de 1 048 576 patients uniques. Dans cet ensemble de donnée, 1 signifie "oui" et 2 signifie "non".

Dans la suite de notre document, nous ne considérons que les individus ayant une classification du test covid égal à 1. Ceci nous permet d'avoir un échantillon homogène d'individus ayant le même degré d'infection au covid. Notre question de recherche étant de déterminer s'il existe un lien de causalité entre le fait de rester à l'hôpital et la mortalité, notre échantillon est constitué d'individus pour lesquelles nous avons d'informations sur leur état de santé en générale, sur leur test covid, s'ils sont restés à l'hôpital ou pas et s'ils sont décédés. Etant donné que nos analyses seront effectuées à partir de modèle logit/probit, nous procédons à un recodage des variables, 1 signifie "oui" et 0 signifie "non". Après avoir supprimé tous les individus dont la classification du test covid est différent de 1 (niveau chronique le plus élevé), nous obtenons un jeu de données qui contient 8 601 observations dont 2 202 observations correspondent aux individus qui sont restés à l'hôpital. Le tableau 1 résume nous sélection d'échantillon ainsi que le nombre de décès par catégorie d'individus.

TABLE 1 – Tableau croisé patient type et décès de l'échantillon sélectionné

Patient type	Décès		Total
	Non (0)	Oui (1)	
Maison (0)	6319	80	6399
Hôpital (1)	1327	875	2202
Total	7646	955	8601

Lorsque nous examinons notre échantillon, nous remarquons que 875 individus parmi ceux qui sont restés à l'hôpital sont décédés soit environ 40% de l'effectif de ceux qui sont restés à l'hôpital, tandis que parmi les individus qui sont rentrés à la maison, le taux de décès est vraiment insignifiant soit 80 individus sur 6 399. Il est à noter que ces résultats descriptifs doivent être pris qu'avec précaution parce qu'ils ne tiennent pas compte du biais de sélection potentiel, à savoir que les individus ayant le covid et qui sont rentrés chez eux diffèrent de ceux qui sont restés à l'hôpital. Il se peut qu'en fonction de la région, des conditions de vie, de la profession, ces tendances statistiques diffèrent. Mais nous ne disposons pas de ces informations sur les individus.

Le tableau 2 résume les informations sur les variables explicatives que nous avons choisies pour notre modèle selon le type de patient. Nous observons des différences importantes en moyenne entre les deux sous-échantillons. Cette différence explique également notre choix pour l'utilisation de la méthode d'appariement par score de propension pour contrôler les différences entre les deux sous-échantillons d'individus afin de déterminer un impact potentiel du fait d'avoir le covid et de rester à l'hôpital sur la mortalité.

TABLE 2 – Statistique sommaire des variables par patient type

Variables	Individus qui sont rentrés à la maison (0)	Individus qui sont restés à l'hôpital (1)
died	1, 2	39, 7
sex	50	36
pneumonia	6, 1	39, 7
age	40	55
pregnant	1, 1	0, 5
diabetes	8, 4	26, 8
copd	0, 5	3, 1
asthma	2, 4	1, 3
inmsupr	0, 6	1, 6
hypertension	11, 5	27, 2
otherdisease	0, 6	3, 7
cardiovasculaire	1, 4	3, 7
obesity	14, 1	15, 7
renalchronique	0, 7	3, 3
tobacco	8, 7	7, 7

De ce tableau, on remarque que parmi ceux qui sont rentrés à la maison, 1,2% sont décédés, 50% sont des femmes, 11,5% sont atteints d'hypertension, 14,1% ont l'obésité et leur âge moyen est de 40ans. La majorité de ceux qui sont restés à l'hôpital ont la pneumonie. 36% sont des femmes et leur âge moyen est de 55ans. 26,8% ont le diabète et 27,2% l'hypertension. Enfin, soulignons qu'il y a beaucoup de différence entre nos deux sous échantillons qui seront source de biais si nous n'utilisons pas une méthode d'appariement pour contrôler les déterminants de l'hospitalisation.

3 Methodology

Pour évaluer l'effet causal du fait de rester à l'hôpital lorsqu'on a le covid sur la mortalité, il est important de pouvoir s'assurer que les individus ont le même état de santé. C'est pour cela nous sélectionnons un échantillon d'individus ayant le même niveau de classification du test covid. Mais comme nous l'avons vu précédemment, les caractéristiques de ceux qui sont restés à l'hôpital après le test peuvent différer considérablement de ceux qui sont rentrés chez eux. Pour contrôler le biais de sélection qui peut être dû aux caractéristiques inobservables, nous utilisons une méthode d'appariement par score de propension. C'est-à-dire que chaque individu qui est resté à l'hôpital est apparié à un ensemble d'individus vivant à domicile avec des caractéristiques similaires. Ceci nous permet d'avoir un contrefactuel par rapport auquel nous pouvons mesurer l'effet de rester à l'hôpital sur la mortalité.

D'après la terminologie de la méthode d'appariement, le fait de rester à l'hôpital est assimilé au fait d'être traité. Il revient donc de trouver des individus qui sont similaire aux individus traités en termes de caractéristiques avant le traitement et qui ne sont pas restés à l'hôpital. L'objectif c'est de pouvoir appairer des individus de deux groupes ayant les mêmes caractéristiques afin que les différences de résultats de ces paires appariés soient liées au traitement, c'est-à-dire au fait d'être resté à l'hôpital. Les individus qui sont rentrés à la maison étant l'unité de contrôle et ceux qui sont restés à la maison, l'unité de contrôle, ce qui compte finalement c'est la mortalité. La méthode d'appariement a pour avantage de ne pas nécessiter la spécification d'une forme fonctionnelle de l'équation du résultat et donc n'est pas susceptible de présenter un biais de mauvaise spécification du modèle. Mieux encore, l'idée de cette méthode est de comparer deux individus qui sur la base des observables, ont une probabilité très proche d'être traités mais l'un a reçu le traitement mais l'autre non.

Score de propension et appariement

Notons que notre analyse est basée sur l'hypothèse d'indépendance conditionnelle (CIA) selon laquelle la mortalité des individus du groupe témoin et du groupe traité est indépendante de là où l'individu décide de rester, une fois que nous contrôlons un ensemble de caractéristiques observables. Pour obtenir les scores de propension, nous faisons des régressions Logit où la variable dépendante est le fait d'être resté à l'hôpital et les variables explicatives sont : sexe, âge, pneumonie, grossesse, diabète, MPOC, asthme, immunodéprimé, hypertension, cardiovasculaire, autre maladie, obésité, maladie rénale chronique, tabac, décès.

Estimation du score de propension

Logistic regression	Number of obs = 8,601			
	LR chi2(14) = 3411.91			
	Prob > chi2 = 0.0000			
	Pseudo R2 = 0.3487			

patient_type	Coefficient	Std. err.	z	P>z
sex	-	.0664925	-7.21	0.000
pneumonia	+	.0747947	38.35	0.000
age	+	.0021731	19.83	0.000
pregnant	+	.3497155	2.43	0.015
diabetes	+	.0936645	3.18	0.001
copd	+	.2821423	2.52	0.012
asthma	-	.2571892	-2.16	0.031
inmsupr	-	.3363953	-0.30	0.766
hypertension	-	.0921627	-1.52	0.128
other_disease	+	.2702055	3.96	0.000
cardiovascular	-	.2291072	-1.47	0.142
obesity	-	.0926904	-3.05	0.002
renal_chronic	+	.2575256	4.34	0.000
tobacco	-	.123601	-3.46	0.001

Cette régression logistique du fait de rester à l'hôpital sur l'ensemble de nos co-variables montrent que le fait d'être une femme, d'être asthmatique, d'être immunodéprimé, d'avoir l'hypertension, d'avoir des maladies cardiovasculaires, d'avoir l'obésité et d'être fumeuse est négativement lié au fait de rester à l'hôpital. Le fait d'avoir la pneumonie, d'être enceinte, d'avoir le diabète, d'avoir une maladie pulmonaire ou autre maladie, d'avoir une maladie rénale chronique augmente la probabilité de rester à l'hôpital.

Enfin, le choix d'un algorithme d'appariement est important dans le cadre de l'analyse par méthode d'appariement. Dans notre travail, nous apparions les individus en utilisant les algorithmes de Nearest neighbor, Radius et Kernel dont nous présentons les résultats dans certains cas mais

notre analyse s'effectuera majoritairement avec les résultats obtenus avec l'algorithme de kernel avec remplacement. Cet algorithme se base sur les moyennes pondérées des individus du groupe de contrôle pour construire le résultat contractuel des individus traités. L'avantage de cette méthode est qu'elle réduit la variance obtenue puisque beaucoup d'informations sont utilisés par rapport aux autres méthodes. Toutefois, il est possible qu'elle utilise des observations qui ne sont pas très bonnes pour l'appariement. C'est donc pour tester la robustesse de la méthode que nous présentons les résultats des autres méthodes afin de comparer.

Analyse de robustesse de l'indicateur

Un indicateur est robuste s'il est peu sensible à la présence d'outliers, le coefficient de corrélation par exemple n'est pas très robuste. Dans le même sens, un modèle est robuste lorsqu'il permet un prolongement des résultats (dans le temps ou pour une population). Dans le but de vérifier la robustesse de notre indicateur (l'ATT), nous ré-estimons les modèles de notre algorithme d'appariement sur des sous-échantillons de notre échantillon d'étude. Nous estimons à nouveau nos modèles Logit sur uniquement les sous-échantillons des hommes, des femmes et des personnes âgées de 85 ans et plus. Les principales différences de moyenne étant au sein de ces sous-groupes, notre indicateur sera donc robuste s'il est peu sensible à ses nouvelles estimations.

4 Résultats

Le tableau 3 nous présentent les estimations de l'effet moyen du traitement sur les personnes traités (ATT) pour notre échantillon considéré. Nos résultats sont présentés selon la méthode d'appariement de Nearest neighbor, Radius et Kernel afin de pouvoir comparer la robustesse de la méthode d'appariement de Kernel. Mais qualitativement les résultats sont presque identiques.

TABLE 3 – Effet moyen du traitement sur les personnes traités (ATT)

Méthodes d'appariement	Traité	Contrôle	ATT	S.E.
Nearest neighbor	2190	6399	0,290 * **	0,01
Radius	2190	6399	0,294 * **	0,01
Kernel	2190	6399	0,290 * **	0,01

Note : *** signifie statistiquement significatif à 1%

Dans l'ensemble, les résultats sont similaires. L'ATT est stable suivant les 3 algorithmes. L'effet de rester à l'hôpital sur la mortalité est positif et significatif à 1%. Les individus qui sont restés à l'hôpital ont environ 30 points de pourcentage de plus de risque de mourir du covid que ceux qui sont rentrés à la maison. Il existe donc un lien causal entre le fait de rester à l'hôpital et mourir

lorsqu'on est diagnostiqué d'un niveau chronique de covid. Ce lien peut s'expliquer par plusieurs facteurs qui dépendent principalement du contexte général de la crise covid. L'état de panique généralisée, la très forte hospitalisation et le nombre de décès excessif sont des facteurs qui ont dû avoir un impact sur l'état psychologique des patients qui sont restés à l'hôpital et également sur les agents de santé qui étaient les plus sollicités durant la crise.

Cependant, il convient de souligner que ces résultats doivent être pris avec attention car il ne faut pas perdre de vue que nous comparons des individus à caractéristiques individuelles différentes (âge, sexe). Nous ne pouvons donc pas assurer à ce stade la stabilité de notre indicateur.

Analyse de robustesse de l'ATT

Afin de tester la robustesse de notre ATT, nous avons effectué quelques modifications dans notre échantillon afin de voir comment l'ATT varie. Il est connu de tous que la pandémie du covid-19 n'a épargné personne, aucun peuple, aucune ethnie, aucune couleur de peau, aucun sexe, enfant, adolescent, jeune, vieux n'a été épargné. Alors notre premier test a été de reprendre notre analyse sur les sous-échantillons des hommes et des femmes, et de voir l'impact sur l'ATT. Le tableau 4 nous présente les résultats de l'estimation de l'ATT sur les hommes et sur les femmes. Les résultats sont présentés à partir de la méthode d'appariement de Kernel.

TABLE 4 – ATT sur les échantillons des hommes et des femmes

Echantillon	Traité	Contrôle	ATT	S.E.
Sur les hommes	1408	3179	0,317 * **	0,01
Sur les femmes	794	3220	0,240 * **	0,01

Note : *** signifie statistiquement significatif à 1%

Nous remarquons que l'ATT reste toujours positif et significatif dans les deux sous-échantillons à 1%. Bien qu'il ne soit pas stable dans les deux sous-échantillons, il reste assez similaire à ce que nous avons trouvé précédemment. Les hommes qui sont restés à l'hôpital ont environ 31 points de pourcentage de plus de risque de mourir du covid à un niveau chronique que ceux qui sont rentrés à la maison. Les femmes qui sont restées à l'hôpital ont environ 24 points de pourcentage de plus de risque de mourir du covid à un niveau chronique que celles qui sont rentrés à la maison.

Une autre question que nous nous sommes posés après ce premier test, c'est de chercher à vérifier si l'ATT est homogène lorsqu'on se base sur les personnes les plus âgées qui ont plus de chances de mourir par rapport aux individus. C'est-à-dire nous constituons à nouveaux deux sous-échantillons. L'un constitué uniquement des personnes âgées de 85 + et l'autre constituée des moins de 85ans. Les résultats de l'estimations à partir de la méthode d'estimation de Kernel sont présentés dans le tableau 5.

TABLE 5 – ATT par groupe d'âge

Echantillon	Traité	Contrôle	ATT	S.E.
Moins de 85ans	6371	2149	0,288 * **	0,01
85ans et plus	49	36	0,337 * **	0,12

Note : *** signifie statistiquement significatif à 1%

Nous remarquons que dans ce cas également l'ATT est positif et significatif dans les deux sous-échantillons. Dans le sous-échantillon des moins de 85ans, le fait de rester à l'hôpital à la suite d'un diagnostic chronique augmente la probabilité de mourir d'environ 28 point de pourcentage. Dans le sous-échantillon des plus de 85ans, le fait de rester à l'hôpital à la suite d'un diagnostic chronique augmente la probabilité de mourir d'environ 33 point de pourcentage.

5 Conclusion

La crise sanitaire du Covid a eu un impact social et économique très important dans tous les pays. La vulnérabilité de tous s'est fait ressentir peu importe l'âge, le sexe, l'ethnie. Un important constat de la pandémie de Covid est le nombre de décès excessif dans les hôpitaux. Les agents de santé subissaient une forte pression en raison du nombre d'hospitalisation journalier excessif et les patients atteints d'un niveau chronique du virus étaient psychologiquement affectés en raison du nombre de décès considérable dans les hôpitaux (ANSM, 2021). Sur la base de ces observations de surmortalité dans les hôpitaux durant la période de forte crise Covid, nous avons voulu vérifier la préexistence de cette tendance.

En utilisant la méthode d'appariement par score de propension, nous avons construit un échantillon dans lequel les individus traités (étant atteint d'un niveau chronique et hospitalisé) et non traités (étant atteint d'un niveau chronique et à domicile) ont des caractéristiques similaires en termes d'âge, de sexe, de caractéristiques socio-démographiques individuelles (fumeur, alcoolique, etc.) et d'état de santé. L'objectif est de déterminer s'il existe un lien de causalité entre le fait d'être à l'hôpital et la probabilité de décès. Nos résultats indiquent que, dans l'ensemble, le fait de rester à l'hôpital à la suite d'un diagnostic chronique de Covid durant la crise augmente la probabilité de mourir plutôt que de rester à domicile. Il convient de noter que ce résultat peut être interprété comme indiquant que, à la suite d'un diagnostic chronique de Covid durant la période de forte crise, recevoir les soins chez soi est une option plus sûre qu'être hospitalisé.

Ce résultat peut être lié à la forte pression subie par les agents de santé et à l'état psychologique affecté des patients chroniquement atteints et hospitalisés. Il ouvre également une piste à des re-

cherches futures sur la prédiction du type de ressources optimales dont un individu pourrait avoir besoin au moment où il est testé positif au Covid.