
Covid-19 : analyse spatiale de l'influence des facteurs socio-économiques sur la prévalence et les conséquences de l'épidémie dans les départements français

Document de Travail
Working Paper
2020-4

Mounir Amdaoud
Giuseppe Arcuri
Nadine Levratto



EconomiX - UMR7235
Université Paris Nanterre
Bâtiment G - Maurice Allais, 200, Avenue de la République
92001 Nanterre cedex

Email : secretariat@economix.fr



Covid-19 : analyse spatiale de l'influence des facteurs socio-économiques sur la prévalence et les conséquences de l'épidémie dans les départements français

Mounir AMDAOU

CEPN, CNRS, université Paris Nord
Chercheur rattaché à EconomiX
Mail : mounir.amdaoud@economix.fr

Giuseppe ARCURI

Université Paris 1 Panthéon Sorbonne, PRISM
EconomiX
Mail : giuseppe.arcuri@univ-paris1.fr

Nadine LEVRATTO (auteur correspondant)

EconomiX, CNRS, université Paris Nanterre
Mail : nadine.levratto@parisnanterre.fr

Résumé : Cet article s'intéresse aux déterminants socio-économiques des hospitalisations et des taux de décès liés à la Covid-19 entre les semaines 9 et 15 d'une part et à la surmortalité observée cette année par rapport aux précédentes, d'autre part. Il propose une approche territoriale de ces questions grâce au recours à des données calculées au niveau des départements français, publiées par Santé France et l'INSEE. L'analyse spatiale exploratoire réalisée révèle d'abord l'hétérogénéité et l'autocorrélation spatiale de la maladie et de ses conséquences. Le recours à des modèles d'économétrie spatiale permet ensuite de mettre en évidence l'influence de la densité démographique, des inégalités sociales, de la part des ouvriers dans la population active et des services d'urgence sur les phénomènes étudiés. Enfin, nous rendons également compte du rôle des effets de débordements entre départements situés à proximité les uns des autres. Ces derniers interviennent davantage dans la diffusion des cas et décès liés à la Covid-19 que dans la surmortalité.

Mots clés : Covid-19, déterminants locaux, analyse spatiale

Codes JEL : C21, I14, R23

1. Introduction

L'apparition, la diffusion et le nombre de décès liés à la Covid-19 en France à partir de la fin du mois de février 2020 a suscité de nombreux commentaires tenant d'abord à leur nombre, au taux de croissance et aux classes d'âge les plus concernées. Très rapidement une autre dimension, régionale, cette fois a été soulignée. L'emploi du terme clusters, pour décrire les foyers d'infection, et le focus sur le Grand Est, les Hauts-de-France et l'Île-de-France qui concentrent le plus grand nombre de cas conduisent à se demander de quels mécanismes répond la répartition spatiale de ce virus.

Les facteurs individuels sont évidemment essentiels pour expliquer la probabilité de contracter la maladie et son issue. Cependant, l'importance des facteurs socioéconomiques locaux comme facteur explicatif de l'état de santé des populations et des taux de mortalité est démontrée par la littérature (voir la revue de littérature de Cambois et Jusot, 2007). Elle est également prise en considération par les Etats comme le rappellent les Objectifs du millénaire pour le développement adoptés en 2000. Ces derniers reposent sur l'idée que la lutte contre la pauvreté, à côté de l'amélioration de la médecine et des systèmes de santé, a des conséquences sur la mortalité, maternelle et infantile notamment. La relation entre la situation économique et sociale et la santé est confirmée pour les pays en développement et les pays développés (Link et Phelan, 1995), dont la France (Ghosn et al., 2012).

Cet article s'intéresse à la répartition spatiale du nombre de cas de Covid-19 nécessitant une hospitalisation et du nombre de décès liés à ce virus recensés dans les départements français. Les données utilisées proviennent de l'INSEE et du ministère de la santé. Les premières analyses à ce sujet ont mis en évidence une forte hétérogénéité spatiale dans répartition du nombre d'individus concernés entre les différents départements de France métropolitaine. C'est pourquoi nous proposons de tester économétriquement l'influence des facteurs socioéconomiques locaux susceptibles d'expliquer la prévalence et la mortalité liée à la maladie et d'analyser le poids respectif des caractéristiques démographiques, des niveaux et de la répartition des revenus et des conditions de vie. Dans la mesure où l'épidémie s'est diffusée sur une partie du territoire et que sa propagation a fait l'objet d'une attention considérable de la part des autorités publiques, nous cherchons également à mettre en évidence la structure géographique des données. Pour ce faire, nous recourons à des techniques d'analyse spatiale permettant de repérer les groupes de départements particulièrement touchés ou épargnés.

L'objectif de cette recherche est donc de contribuer à la compréhension des vecteurs de propagation de l'épidémie et d'en souligner la dimension socio-économique. Elle s'inscrit dans le champ de la géographie de la santé qui a émergé dans les années 1970 et a connu un profond renouvellement dans les années 1990 sous l'influence de travaux de recherche nombreux et variés des domaines de l'épidémiologie, des sciences de l'environnement, des sciences sociales, de la santé publique ou du

management des services dans le but de rendre compte des disparités spatiales en matière d'accès aux soins, d'exposition aux maladies et de morbidité (Rosenberg, 1998 ; Curtis, 2004 ; Prior et al. 2018). Cette recherche complète aussi celles entreprises par les épidémiologistes qui pointent les facteurs individuels tels sur l'âge, l'indice de masse corporelle ou l'état de santé des personnes touchées par la Covid-19 (Onder, Resa et Brusaferro, 2020 ; Wu et McGoogan, 2020 ; Diez et Santos-Burgoa, 2020) comme le recense la base de données Worldometer¹.

La prise en considération d'éléments de contexte dans l'analyse d'une question liée à la santé est nécessaire à plusieurs niveaux. Tout d'abord, parce que les différences d'un territoire à l'autre en matière de santé ou de maladie sont de plus en plus souvent pointées par la littérature et que leur omission conduirait à une compréhension partielle des phénomènes étudiés comme le soulignent Geronimus et al. (1999) ou, plus récemment et à propos de la Covid-19 en Italie, Bayer et Kuhn (2020)². Ces derniers envisagent ainsi que les structures familiales et la présence au sein d'un même logement de familles comptant plusieurs générations diffèrent selon les régions et expliquerait ainsi les écarts géographiques observés. Ensuite, car la responsabilité des différents types de caractéristiques d'un territoire peuvent jouer différemment selon le type d'affection. Ainsi, une importante littérature pointe la responsabilité de la pauvreté dans la prévalence du choléra comme l'ont montré Talavera et Perez (2009) dans une comparaison internationale réalisées à partir de données de la Banque Mondiale ou Olson et al. (2012) pour la tuberculose. La qualité du système de santé peut également expliquer les écarts entre régions (Scott et Coote, 2007) comme c'est par exemple le cas pour le diabète (Scott et al. 2009). Enfin, la démographie intervient comme un facteur de différenciation spatiale en matière de santé et, à ce titre, fait l'objet de fréquentes analyses pour expliquer les écarts internationaux (Hu et Goldman, 1990) et régionaux (Frohlich et Mustard, 1996). La troisième raison qui légitime une approche socio-économique à l'échelle géographique de la Covid-19, est l'organisation du système de santé. A partir de l'entrée en vigueur de la loi relative à la politique de santé publique en 2004, l'échelon local est devenu le niveau pertinent en matière de santé publique. La loi portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires (HPST), du 21 juillet 2009, a créé les Agences régionales de santé (ARS) dans un souci de gestion de la politique de santé au plus près du terrain et a introduit la notion de Contrat local de santé (CLS) afin d'articuler les politiques régionales de santé, l'action des collectivités territoriales et des autres acteurs locaux autour d'un diagnostic partagé. De même, le territoire est présent dans la réorganisation hospitalière avec la création des Groupements hospitaliers de territoires (GHT).

¹ Disponible en ligne : <https://www.worldometers.info/coronavirus/>

² A ce stade, il s'agit de pistes de réflexion qui nécessiteront des analyses approfondies lorsque de nouvelles données seront disponibles. Cette hypothèse et les tests réalisés à son sujet sont discutés par Buonnanno et al. (2020).

Notre démarche empirique repose sur les cartes de diffusion de la Covid-19 par département communiquées par l'INSEE et le Ministère des Solidarités et de la Santé depuis le début du mois de mars 2020 et des commentaires d'experts, analystes et journalistes réalisés à chaud et relayés par la presse. Nous en résumons le contenu. Dès le début, les mouvements des populations aisées des grandes métropoles, notamment Paris, vers leurs résidences secondaires localisées dans les départements de Normandie et de Bretagne et de Vendée, ont été pointés comme facteurs de diffusion de l'épidémie. Plusieurs interventions ont également mis l'accent sur la densité comme facteur explicatif des taux élevés de Coronavirus, en Ile de France notamment, et sur la situation des hôpitaux publics et de la médecine de ville. Enfin, plus récemment, le rôle de la pauvreté a été relevé par la presse pour expliquer le non-respect du confinement, en Seine Saint Denis notamment.

Suivant cette logique, nous mobilisons trois grandes familles de variables locales pouvant potentiellement affecter le taux de malades et le taux de mortalité liés à la Covid-19 : des déterminants économiques, des déterminants démographiques et des déterminants liés au cadre de vie. Afin de rendre compte des effets de diffusion et du rôle des différentes familles de déterminants analysés dans ce processus, nous menons une analyse exploratoire de l'autocorrélation spatiale qui permet de mesurer la dépendance spatiale entre les valeurs d'une même variable en différents endroits de l'espace. En effet, les données par département, comme nombre de variables pour lesquelles on dispose d'informations géolocalisées, se caractérisent par des dépendances spatiales qui sont d'autant plus fortes que les localisations sont plus proches. Tenir compte de cette particularité permet donc de prendre en compte les interactions et les externalités spatiales dans l'analyse des phénomènes observés.

Les résultats obtenus mettent en évidence le rôle des facteurs socio-économiques dans les manifestations de la Covid-19 quel que soit l'indicateur obtenu. La densité, la part d'ouvriers dans la population active et les inégalités de revenus influencent positivement le taux d'hospitalisation, le taux de décès et la surmortalité. Le nombre de services d'urgence corrigé de la taille du département exerce en revanche une influence négative sur les trois variables expliquées retenues. Enfin, aucune influence significative de la part des résidences secondaires n'est mise en évidence, ce qui nous conduit à relativiser l'importance des mouvements de population intervenus au moment du confinement.

Le reste de l'article s'organise de la façon suivante. La section 2 met en évidence l'hétérogénéité et le *clustering* spatial de la Covid-19. La section 3 présente les données et les variables explicatives introduites dans l'analyse. La section 4 expose les techniques d'estimation retenues et la section 5 présente et commente les résultats obtenus. Enfin, la section 6 conclut.

2. Hétérogénéité spatiale des cas de Covid-19 et de la mortalité associée

Cette section rend compte de l'hétérogénéité spatiale de l'épidémie de Covid-19 et des taux de mortalité associés. Elle repose sur des données au niveau des départements récemment mises à disposition par le ministère des Solidarités et de la Santé³ et l'INSEE⁴. Notre analyse couvre les 96 départements de France métropolitaine. Les départements d'outre-mer sont exclus de l'analyse car le nombre de cas recensés restait trop faible au moment où les données ont été collectées, de l'absence de certaines données mobilisées dans cette recherche et de la discontinuité géographique. Pour chaque département, trois indicateurs d'intensité de l'épidémie ont été calculés :

- i) un indicateur classique de prévalence mesuré par le nombre d'hospitalisations liés à la Covid-19. Il est défini par la somme des nouvelles hospitalisations entre le 19 mars et le 13 avril rapporté au nombre d'habitants du département pour 1000 habitants ;
- ii) le taux de mortalité mesuré par le nombre de décès dus à la Covid-19 à l'hôpital⁵ recensés le 12 avril rapporté au nombre d'habitants du département pour 1000 habitants ;
- iii) un taux de surmortalité mesuré par la différence entre le nombre total de décès par département au 30 mars 2020⁶ et la moyenne des décès recensés à la même date en 2018 et 2019 rapportée à cette dernière qui permet de tenir compte des décès liés au virus en dehors du milieu hospitalier, dans les EHPAD notamment⁷.

Cette dernière variable a été construite par référence à la moyenne des deux années qui précèdent en raison de la durée exceptionnellement longue de l'épisode de grippe en 2018 et de sa relative faiblesse en 2019. Ainsi au niveau national, le nombre de décès totaux survenus entre le 1er et le 30 mars 2020 est supérieur à celui enregistré sur la même période en 2019 : 57 441 décès ont été enregistrés en 2020 en France contre 52 011 en 2019. Ce nombre reste néanmoins inférieur au nombre des décès (58 641) enregistrés sur la même période en 2018. Nous n'avons pas retenu le taux de létalité mesuré par le nombre de décès rapporté au nombre de cas déclarés en raison des débats portant sur la fiabilité de ce ratio en l'absence de tests à grande échelle. Au 13 avril 2020, on dénombrait 70 265 passages à l'hôpital enregistrés depuis le 1^{er} mars sur la France entière alors que les données actualisées sur les décès en milieu hospitalier en reportent 9 234 au 12 avril.

Les cartes reportées sur les figures 1, 2 et 3 rendent compte de l'hétérogénéité spatiale, très tôt décelée par le secteur médical, de la maladie. Les départements les plus touchées, quel que soit l'indicateur retenu, sont ceux de la région Grand Est, des Hauts-de-France et de l'Ile-de-France. On note

³ Les fichiers sont regroupés sur le site data.gouv.fr.

⁴ Les données sont disponibles à l'adresse suivante : <https://www.insee.fr/fr/information/4470857>

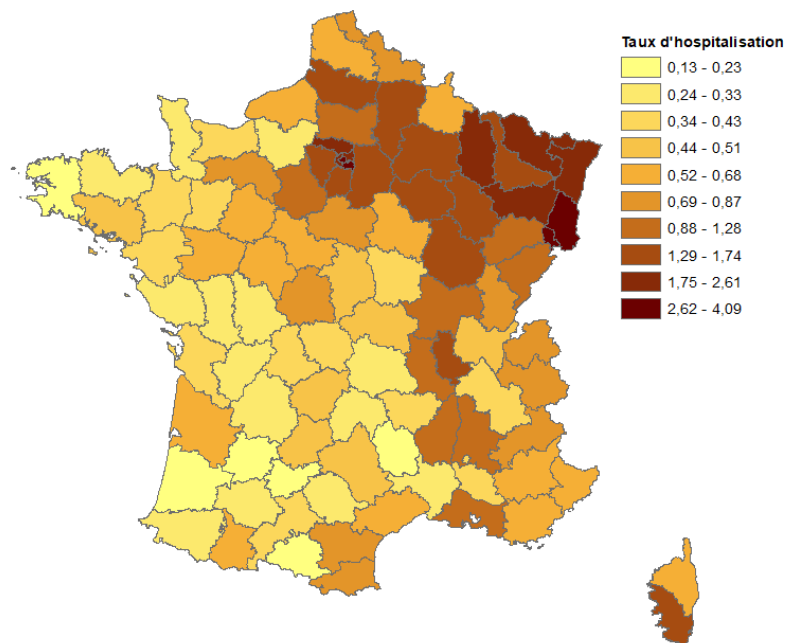
⁵ Le périmètre hospitalier pour le nombre de décès est imposé par la comptabilisation des décès en EHPAD à partir du mois d'avril seulement.

⁶ Les données ont été enregistrées à la date du 10 avril 2020.

⁷ Les données sur ces derniers ont commencées à être intégrées aux statistiques sur les décès à partir du début du mois d'avril.

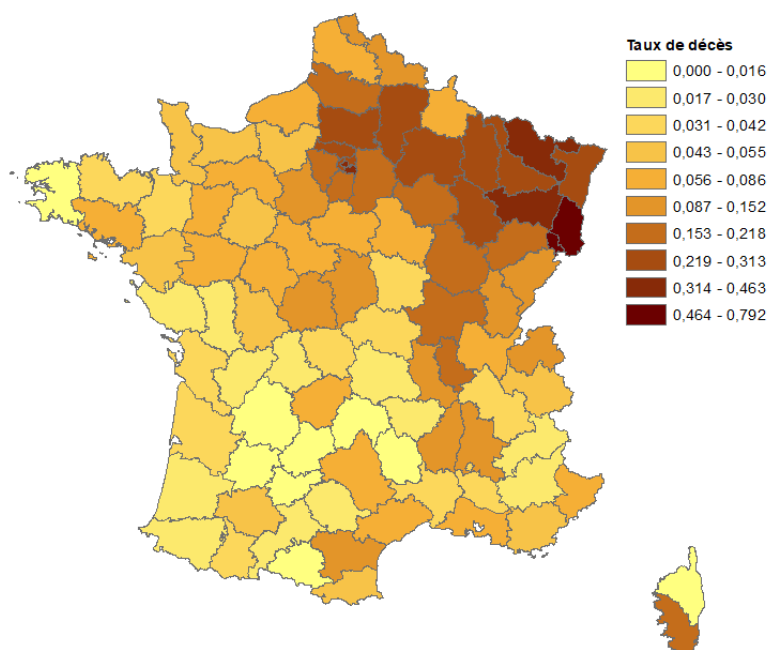
également de plus forts taux d'hospitalisation et de mortalité le long de la vallée du Rhône. En revanche, la moitié ouest du pays apparaît relativement moins touchée par la Covid-19, cette moindre exposition étant particulièrement visible dans le quart sud-ouest.

Figure 1 : Taux d'hospitalisation pour la Covid-19 par département (par décile)



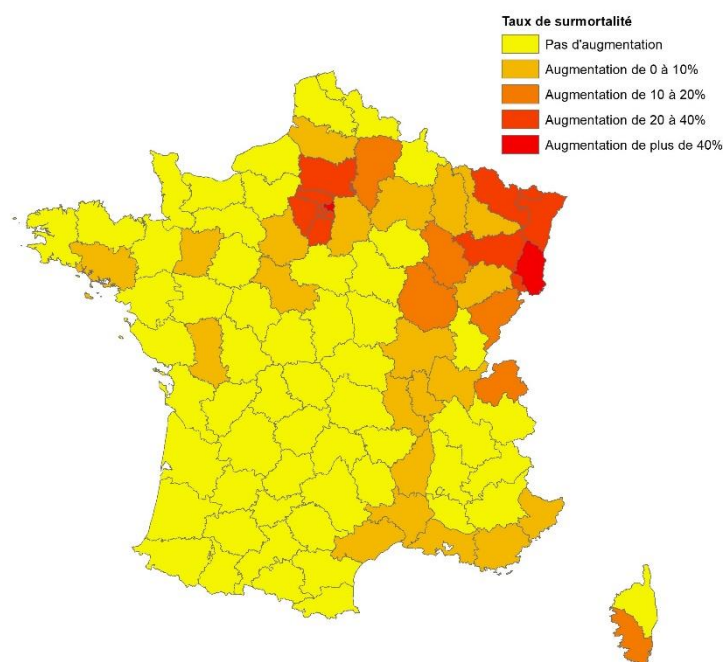
Source : Ministère des solidarités et de la santé. Calculs et cartographie des auteurs

Figure 2 : Taux de décès dus à la Covid-19 en milieu hospitalier par département (par décile)



Source : Ministère des solidarités et de la santé. Calculs et cartographie des auteurs

Figure 3 : Taux de surmortalité par département au 30 mars 2020
par rapport à la moyenne de 2018 et 2019



Source : INSEE. Calculs et cartographie des auteurs

La polarisation dans l'espace des indicateurs analysés peut résulter d'un effet de contagion propageant la maladie d'un département à l'autre. En présence de ce type de regroupement spatial des données, l'utilisation des méthodes habituelles d'analyse s'accompagne d'un risque de biais. En effet, si le phénomène observé dans le département i est influencé par ce qui se passe dans le département j , la condition de normalité des résidus n'est plus respectée. Afin de tester l'existence d'un phénomène de clustering spatial des données, nous appliquons la méthode ESDA (Exploratory Spatial Data Analysis), préconisée pour analyser les données locales (Baumont et al., 2006; Guillain et Le Gallo, 2010).

La détermination des interactions entre les différents départements repose sur la définition d'une matrice de poids spatial (Florax et Nijkamp, 2004). La matrice spatiale W ici utilisée est celle dite de distance inverse⁸. La pondération inverse à la distance est une méthode d'interpolation spatiale classique établie selon un processus permettant d'assigner une valeur à tout point d'un espace à partir d'un semis de points connus, ici les centroïdes des départements.

L'autocorrélation spatiale est mesurée par la statistique I de Moran (Upton et Fingleton, 1985 ; Haining, 1990). Elle représente pour chaque variable x le degré d'association linéaire entre sa valeur à

⁸ Les matrices de distance inverse prennent en compte la relation décroissante entre l'intensité de la relation et la distance. Elles prennent la valeur 1 si $d < d_0$, $1/d^2$, e^{-2d} , etc.

une certaine localisation et la moyenne pondérée spatialement de ses voisins. Cette statistique s'écrit de la façon suivante :

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} (y_{it} - \bar{y}_t) * (y_{jt} - \bar{y}_t)}{\sum_i (y_{it} - \bar{y}_t)^2}$$

dans laquelle I est l'indice d'autocorrélation spatiale de Moran, w_{ij} , est un élément de la matrice de poids spatial W . Les valeurs varient entre -1 (dispersion parfaite) à +1 (corrélation parfaite). Une valeur nulle est significative d'un modèle spatial parfaitement aléatoire. Les valeurs négatives (respectivement positives) de l'indice indiquent une autocorrélation spatiale négative (respectivement positive).

Le Tableau 1 présente les valeurs de la statistique I de Moran des trois indicateurs d'intérêt dans les 96 départements de France métropolitaine aux dates retenues pour cette recherche.

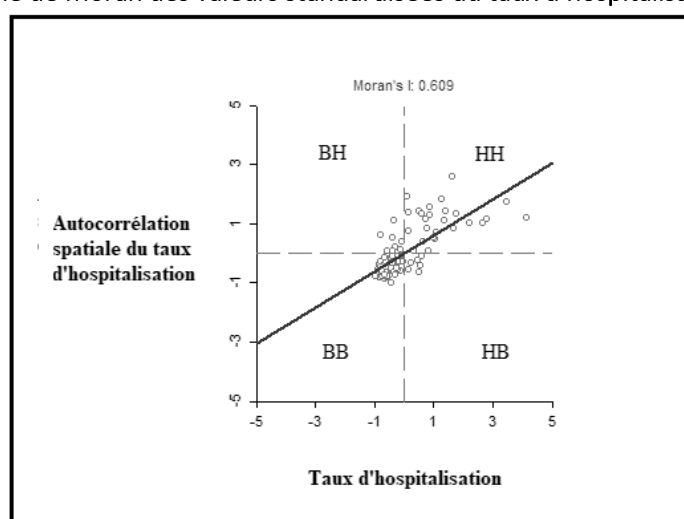
Tableau 1 : Statistique I de Moran

Indicateur	I de Moran	Moyenne	Ecart-type	Valeur standardisée	P-value
Taux d'hospitalisation	0.609	-0.0103	0.0756	8.1968	0.0001
Taux de décès en milieu hospitalier	0.642	-0.0102	0.0738	8.8402	0.0001
Taux de surmortalité	0.560	-0.0112	0.0731	7.7982	0.0001

Source : Santé Publique France et INSEE. Calculs des auteurs.

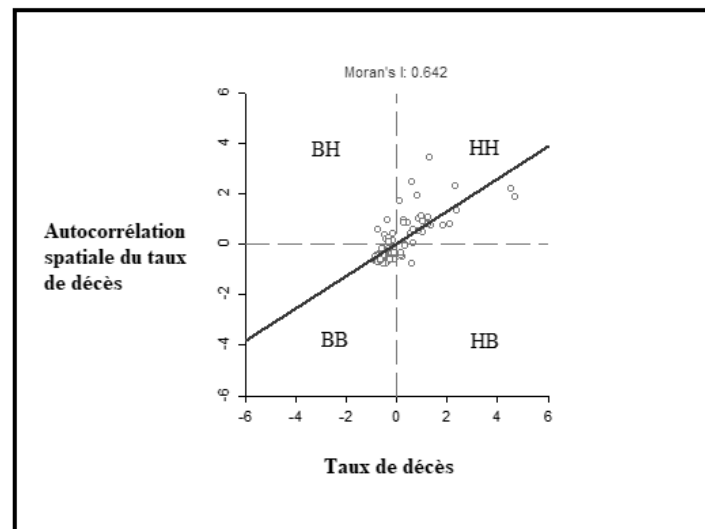
Les valeurs des coefficients obtenues 0.609, 0.642 et 0.560, indiquent la présence d'une importante autocorrélation spatiale positive. Ainsi, les départements qui présentent des taux d'hospitalisation, de mortalité et de surmortalité élevés (respectivement faibles) ont tendance à se trouver à proximité de départements présentant des valeurs des taux similaires. Les diagrammes des figures 4, 5 et 6 rendent compte de ces regroupements spatiaux.

Figure 4 : Diagramme de Moran des valeurs standardisées du taux d'hospitalisation lié à la Covid-19



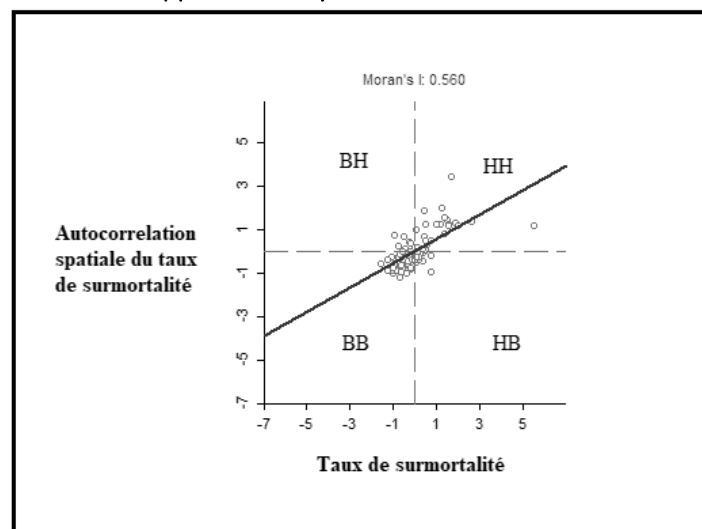
Source : Santé Publique France. Calculs des auteurs

Figure 5 : Diagramme de Moran des valeurs standardisées du taux de décès en milieu hospitalier lié à la Covid-19



Source : Santé Publique France. Calculs des auteurs

Figure 6 : Diagramme de Moran des valeurs standardisées du taux de surmortalité au 30 mars 2020 par rapport à la moyenne de 2018 et 2019



Source : INSEE. Calculs des auteurs

Les LISA (Local Indicator of Spatial Association) proposés par Anselin (1995) viennent compléter l'indice I de Moran en offrant une vision plus précise du *clustering* local. Ils permettent de tester l'hypothèse de distribution aléatoire en comparant les valeurs de chaque localisation spécifique avec les valeurs dans les localisations voisines.

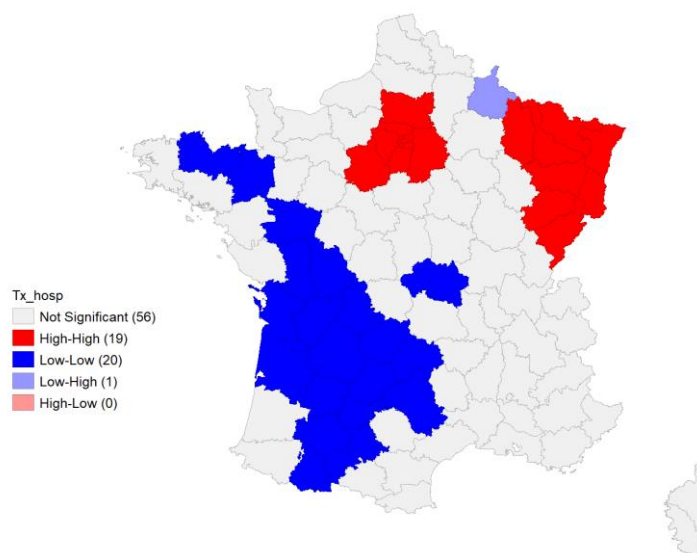
Cinq possibilités sont identifiées :

- Les regroupements de type haut-haut (HH), ou points chauds, où l'autocorrélation spatiale est positive tels que les unités spatiales présentant une valeur élevée de la variable sont entourées d'unités spatiales similaires,

- Les regroupements de type bas-bas (BB), ou points froids, où l'autocorrélation spatiale est également positive mais, qui à la différence des précédents sont tels que les unités spatiales présentant une valeur faible de la variable sont entourées d'unités spatiales similaires,
- Les regroupements atypiques de type haut-bas (HB), où l'autocorrélation spatiale est négative tels que les unités spatiales présentant une valeur élevée de la variable sont entourées d'unités spatiales dans lesquelles la variable prend une valeur faible
- Les regroupements atypiques de type bas-haut (BH), où l'autocorrélation spatiale est également négative mais que sont tels que les unités spatiales présentant une valeur faible de la variable sont entourées d'unités spatiales dans lesquelles la variable prend une valeur forte,
- Les unités spatiales pour lesquelles aucune autocorrélation spatiale significative n'est repérée.

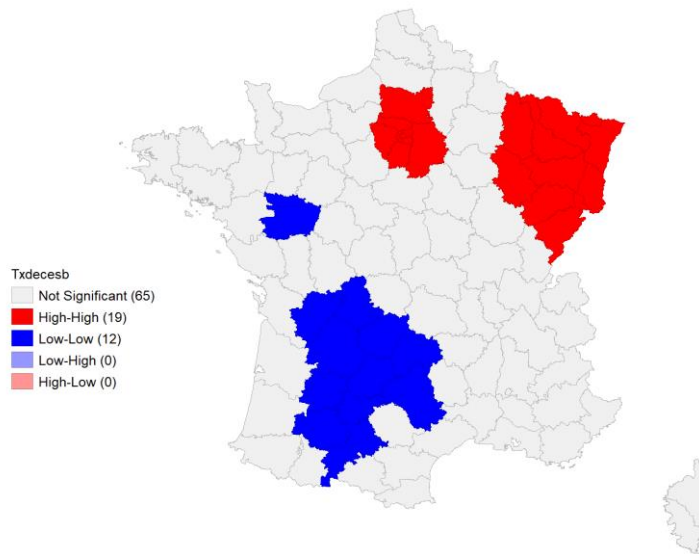
Les cartes des figures 7, 8 et 9 représentent les autocorrélations spatiales des taux d'hospitalisation, de mortalité et de surmortalité par département données par les statistiques LISA.

Figure 7 : LISA-Taux d'hospitalisation de la Covid-19 par département



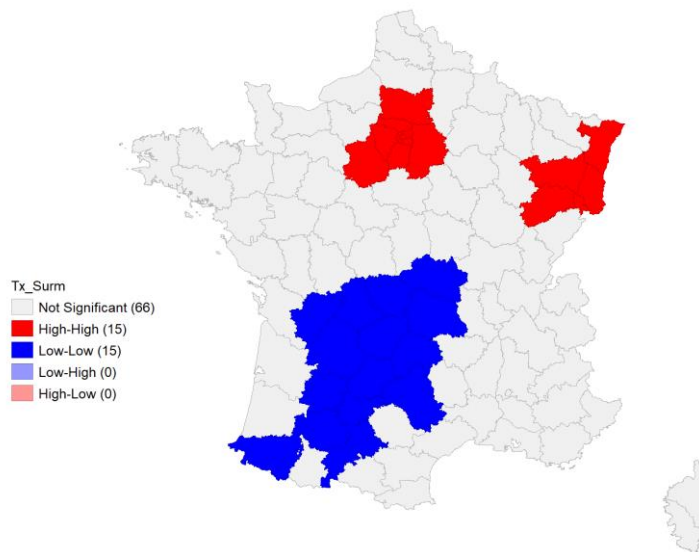
Source : Santé Publique France. Calculs des auteurs.

Figure 8 : LISA-Taux de décès liés à la Covid-19 par département



Source : Santé Publique France. Calculs des auteurs.

Figure 9 : LISA-Taux de surmortalité par département



Source : INSEE. Calculs des auteurs.

La carte des LISA pour les taux d'hospitalisation rend compte de la présence de deux clusters de type HH situés dans l'est de la France et dans le bassin parisien. Un grand quart sud-ouest et une partie de la Bretagne sont au contraire caractérisés par des taux d'hospitalisation significativement plus faibles que sur le reste du territoire. Ces départements forment un grand cluster de type BB. A l'exception du département des Ardennes caractérisé par une autocorrélation négative rapportée au cluster du nord-est, aucune autocorrélation spatiale n'est décelée sur le reste du pays. La carte des taux de mortalité présente une structure similaire pour le cluster de type HH mais le cluster de type BB est de moindre

ampleur. Il touche seulement le sud-ouest tandis que la Bretagne et le Centre deviennent non significatifs. Cette même structure d'ensemble est observée pour les LISA de la surmortalité mais on note que les départements composant le cluster HH à l'est sont, dans ce cas, moins nombreux.

3. Identification des variables explicatives

Notre base de données est constituée des 96 départements de France métropolitaine. Dans les analyses qui suivent, nous utilisons soit des ratios pondérés par la population pour tenir compte des différences de taille entre les départements⁹, soit des distributions pour refléter les disparités pouvant exister au sein d'un même département.

Nous distinguons trois natures de variables définies à l'échelle du département susceptibles d'expliquer l'hétérogénéité spatiale des cas et des décès liés à la Covid-19 : des déterminants démographiques, des déterminants économiques et des déterminants liés au cadre de vie. La plupart des informations sont fournies par l'INSEE. Nous avons considéré la dernière année disponible au moment de la rédaction de cet article qui, selon les cas est 2016 ou 2018.

Afin de comprendre l'intensité de la Covid-19 et ses conséquences sur la mortalité au sein de chaque département d'une part et sa diffusion spatiale de l'autre, nous intégrons à notre modèle les variables démographiques suivantes : la densité de population au km², la part des ouvriers dans la population totale et le taux d'inactifs. La densité est prise en compte en tant que facteur aggravant de la contagion (Tarwater et Martin, 2001), la part des ouvriers plus exposée à la continuité du travail (DARES, 2020) et donc à la maladie ainsi que le taux d'inactifs qui capte les personnes âgées, plus affectées par la maladie, et les étudiants scolarisés, considérés comme des vecteurs de propagation de la Covid-19.

Nous les complétons par une variable économique qui est l'écart interdécile de la distribution des revenus. Elle est introduite en tant que proxy des inégalités économiques qui caractérisent les départements et exercent une influence sur l'émergence et la propagation des maladies (Farmer, 1996, Mamelund, 2017) en raison des différences d'accès aux aménités selon le type de population. Trois variables sur le cadre de vie sont considérées. D'abord, la part des résidences secondaires introduite en raison des débats publics sur le rôle des mouvements de population des grandes métropoles, notamment, vers les départements plus touristiques suite à l'annonce du confinement. Ensuite le nombre de services d'urgences par département pour rendre compte de la capacité d'accueil des malades et une variable binaire distinguant les départements urbains et ruraux construite à partir du classement fourni par l'Assemblée des départements de France (ADF).

⁹ En 2016, on dénombrait 76 422 habitants en Lozère et 2 603 723 dans le Nord

Le tableau 2 ci-dessous résume la définition des différentes variables intégrées dans nos modèles. Les statistiques descriptives et la matrice des corrélations se trouvent en annexe.

Tableau 2 : Définition, année de référence et source des variables

Libellé	Définition	Date/Année	Source
Variables expliqués			
Taux d'hospitalisation	(Nombre d'hospitalisation Covid-19/population) *1000	19 Mars -12 Avril 2020	Santé Publique France
Taux de décès	(Nombre de décès Covid19 /population) *1000	19 Mars -13 Avril 2020	Santé Publique France
Taux de surmortalité	Taux d'évolution des décès cumulés au 30 mars 2020 par rapport à la moyenne des décès cumulés entre 2018-2019 à la même date.		INSEE
Variables explicatives			
Densité de population	Log (Nombre d'habitants/ km ²)	2016	INSEE
Taux des ouvriers	Part de la population de 15 ans ou plus selon la CSP ¹⁰ Ouvrier (2016)	2016	INSEE
Taux des inactifs	Part de la population de 15 ans ou plus selon la CSP Autres + personnes sans activité professionnelle - (2016)	2016	INSEE
Ecart niveau de vie	Rapport interdécile des revenus (9 ^e décile/1 ^{er} décile)	2016	INSEE
Nombre de services d'urgence	Log (Nombre Service d'urgences)	2018	INSEE
Part des résidences secondaires	Part des résidences secondaires (y compris les logements occasionnels) dans le total des logements	2016	INSEE
Département rural	Variable dichotomique prenant la valeur 0 si le département est rural ou composé de petites villes et 1 sinon.		ADF

4. Modèles mobilisés

La spécification économétrique considérée dans cette recherche prend comme point de départ le modèle de régression linéaire des moindres carrés ordinaires (MCO) :

$$Y = X\beta + \varepsilon \quad (1)$$

Y représente la variable dépendante (taux d'hospitalisation, taux de décès, taux de surmortalité). X représente les variables explicatives utilisées, β est le vecteur de paramètres à estimer et ε est le terme d'erreur. Lorsqu'un phénomène d'autocorrélation spatiale est ignoré dans la spécification du modèle, mais présent dans le processus générateur des données, les estimateurs des MCO sont biaisés et non convergents.

¹⁰ Catégorie socio-professionnelle

Le modèle spatial autorégressif (SAR) consiste à corriger ce biais en intégrant une “variable endogène décalée” WY dans le modèle (1) et à prendre en compte l'autocorrélation spatiale relative à la variable Y . Le modèle s'écrit comme suit :

$$Y = \rho WY + X\beta + \varepsilon \quad (2)$$

WY est la variable endogène décalée pour la matrice de distance inverse W , ρ est le paramètre autorégressif indiquant l'intensité de l'interaction existante entre les observations de Y . Dans ce modèle, l'observation de Y est en partie expliquée par les valeurs prises par Y dans les départements voisins.

L'introduction de la variable WY dans le modèle (1) est un moyen d'apprécier le degré de dépendance spatiale alors que les autres variables sont contrôlées. Symétriquement, il permet de contrôler la dépendance spatiale pour évaluer l'impact des autres variables explicatives.

Une seconde façon d'incorporer l'autocorrélation spatiale dans les modèles économétriques est le Spatial Error Model (SEM) qui consiste à spécifier un processus de dépendance spatiale des erreurs dans un modèle de régression. Le modèle SEM est défini de la façon suivante :

$$Y = X\beta + \varepsilon \text{ avec } \varepsilon = \lambda W\varepsilon + u \quad (3)$$

Le paramètre λ reflète l'intensité de l'interdépendance entre les résidus de la régression et u est le terme d'erreur. Omettre une autocorrélation spatiale des erreurs produit des estimateurs non biaisés mais inefficients, si bien que l'inférence statistique basée sur les MCO sera biaisée.

Suivant Le Gallo (2002), différentes approches peuvent être utilisées pour le choix des modèles. Nous avons retenu l'approche dite ascendante qui consiste à commencer avec le modèle non spatial. Des tests du multiplicateur de Lagrange (Anselin et al., 1996) permettent ensuite de trancher entre le modèle SAR, SEM ou le modèle non spatial¹¹.

5. Résultats : présentation et commentaires

Les résultats des estimations des déterminants du taux d'hospitalisation, de décès et de surmortalité sont présentés dans les tableaux 3, 4 et 5 en tenant compte des différentes spécifications du modèle présentées (MCO, SAR et SEM). Les tests du multiplicateur de Lagrange ici réalisés montrent une meilleure performance du modèle SAR. Nous commentons donc ce dernier.

¹¹ Le choix entre les trois modèles s'effectue en comparant les niveaux de significativité des tests du multiplicateur de Lagrange (LMError, LMLag) et leurs versions robustes (RLMError et RLMLag). Si LMLag est aussi significatif que LMError et que seul RLMLag est significatif, on inclut une variable endogène décalée et le modèle privilégié sera le SAR. Inversement, si LMError est aussi significatif que LMLag et que seul RLMError est significatif, alors on choisit le modèle avec autocorrélation des erreurs (SEM) (Le Gallo, 2002). Si aucun test n'est significatif le modèle MCO doit être retenu.

Les résultats de l'estimation des déterminants du taux d'hospitalisation sont présentés dans le tableau 3. Ils mettent en évidence le rôle positif de la densité, conforme à celui attendu. L'importance de la densité de population dans la Covid-19 a été fortement discutée dès le début de l'épidémie en Chine. Le rôle de cette variable est d'autant plus important qu'il est lié au nombre de reproduction de base (connu sous l'abréviation R_0) souvent estimé entre 2 et 4 (Roques et al., 2020), reconnu comme très élevé par les épidémiologistes. Le taux d'hospitalisation est également positivement et significativement corrélé à la part des ouvriers dans la population active. Cela confirme le risque d'exposition accru de cette catégorie sociale (conducteurs des systèmes de transport, livreurs, services de voirie et de nettoyage dans la fonction publique territoriale, réparations d'urgence dans les industries de réseaux, etc.) en raison des moindres possibilités de pratique du télétravail et de leur participation plus élevée aux plans de continuité d'activité mis en œuvre dès le début du confinement. Les inégalités sociales approximées par l'écart interdécile des revenus influencent également de manière positive et significative le taux d'hospitalisation. On retrouve ici les résultats de publications antérieures qui montrent que les inégalités sociales sont des déterminants de la santé des populations (Karlsdotter et al. 2012 ; Goldman, 2001) et des facteurs de propagation des épidémies (Tarwater et Martin, 2001 ; Hsueh et al. 2012). Enfin, le nombre de services d'urgences exerce une influence significative négative sur le taux d'hospitalisation, semblant ainsi confirmer qu'une bonne prise en charge médicale permet de réduire le risque d'évolution dangereuse de la maladie¹².

Il est à noter que la part des résidences secondaires n'est pas significative. Elles avaient été soupçonnées d'avoir aggravé la propagation de l'épidémie au cours des jours ayant suivi l'annonce du confinement en attirant des ménages en provenance de la région parisienne, très touchée par l'épidémie.

L'autocorrélation spatiale, identifiée dans l'ESDA, se retrouve dans les estimations. La valeur du coefficient ρ est de 0,505 et significative à 1%. Ce résultat traduit un effet de diffusion entre les départements localisés à proximité les uns des autres¹³.

¹² Il n'a pas été possible d'introduire le nombre de médecins pour 1000 habitants en raison de la forte corrélation (0,87) entre cette variable et le nombre de services d'urgences.

¹³ Les résultats obtenus sont robustes au type de matrice spatiale utilisés. Nous les avons calculés avec une matrice de contiguïté d'ordre 1 de type reine. Nous les tenons à la disposition des lecteur sur simple demande.

Tableau 3 : Résultats des estimations du taux d'hospitalisation

Modèle	MCO	SAR	SEM
Densité de population	0.310** (0.130)	0.186** (0.0781)	0.311*** (0.0987)
Part d'ouvriers	13.16** (5.114)	5.683* (3.285)	-0.0368 (4.251)
Part d'inactifs	7.293* (3.910)	3.155 (2.967)	-0.426 (4.242)
Ecart de niveau de vie	0.722** (0.283)	0.439*** (0.170)	0.250 (0.184)
Nombre de services d'urgence	-0.469** (0.200)	-0.322*** (0.0915)	-0.335*** (0.0911)
Part des résidences secondaires	-0.914 (0.759)	-0.325 (0.604)	-0.121 (0.796)
Département rural	-0.0379 (0.160)	0.00759 (0.113)	-0.0127 (0.107)
Constant	-4.244*** (1.174)	-2.259*** (0.802)	-0.697 (1.080)
ρ		0.505*** (0.0760)	
λ			0.675*** (0.0801)
Observations	96	96	96
R ²	0.564	0.608	0.442
Log Likelihood	-71.323	-55.081	-57.187
AIC	158.646	130.162	134.374
Tests			
I de Moran	4.748***		
LMError	15.962***		
RLMError	0.491		
LMLag	31.309***		
RLMLag	15.839***		

Ecart-types entre parenthèses

***Significatif à 1%, **Significatif à 5%, * Significatif à 10%

Les résultats des estimations du taux de décès liés à la Covid-19 en milieu hospitalier (Tableau 4) vont dans le même sens que les précédents. Ils mettent en évidence l'influence positive significative de la densité de population, de la part d'ouvriers dans la population active et des écarts de niveau de vie. Ils révèlent aussi l'influence significative négative du nombre de services d'urgences. Comme précédemment, nos résultats ne montrent pas d'influence de la part d'inactifs, plus souvent confinés que les actifs, du fait de la fermeture des écoles et des règles de prudence renforcée appliquées aux personnes âgées. L'absence de significativité de la part des résidences secondaires est de nouveau observée.

Les taux de décès sont également autocorrélés spatialement comme l'indique la valeur de 0,550 prise par le coefficient ρ .

Tableau 4 : Résultats des estimations du taux de décès en milieu hospitalier

Modèle	MCO	SAR	SEM
Densité de population	0.0532** (0.0259)	0.0324** (0.0149)	0.0544*** (0.0191)
Part d'ouvriers	3.342*** (1.170)	1.631** (0.638)	1.230 (0.797)
Part d'inactifs	1.453* (0.830)	0.470 (0.574)	-0.188 (0.823)
Ecart de niveau de vie	0.129** (0.0626)	0.0764** (0.0325)	0.0590* (0.0351)
Nombre de services d'urgence	-0.0838* (0.0445)	-0.0532*** (0.0176)	-0.0537*** (0.0177)
Part des résidences secondaires	-0.182 (0.115)	-0.0478 (0.116)	-0.0666 (0.153)
Département rural	-0.0196 (0.0334)	-0.00959 (0.0217)	-0.0110 (0.0209)
Constant	-0.905*** (0.270)	-0.477*** (0.155)	-0.329 (0.203)
ρ		0.544*** (0.0738)	
λ			0.662*** (0.0770)
Observations	96	96	96
R ²	0.499	0.550	0.418
Log Likelihood	83.526	102.422	100.523
AIC	-151.051	-184.844	-181.047
Tests			
I de Moran	5.791***		
LMError	24.867***		
RLMError	0.111		
LMlag	38.431***		
RLMLag	13.676***		

Ecart-types entre parenthèses

***Significatif à 1%, **Significatif à 5%, * Significatif à 10%

L'estimation des taux de surmortalité (tableau 5), qui ne sont pas exclusivement liés à l'épidémie, permet cependant de capter les décès à les EHPAD et ceux des personnes non testées mais décédées de la Covid-19. Les résultats obtenus sont similaires aux précédents pour les variables démographiques et la variable économique. L'influence négative des services d'urgence disparaît dès que l'ensemble des décès (environ 600 000 par an selon l'INED) est pris en considération. Leur répartition peut expliquer cette absence de résultats. En effet, en 2016, 59,2 % des décès ont eu lieu dans des établissements de santé (hôpital ou clinique)¹⁴ (Bellamy, 2017) si bien que pour nombre d'entre eux, les services d'urgences ne sont pas un point de passage contrairement à ce qui se produit pour la Covid-19.

La valeur prise par le coefficient ρ est égale à 0.384. L'effet de débordement, observé pour les variables expliquées précédente se retrouve de nouveau ici, mais à un degré moindre.

¹⁴ Le reste se répartit entre 26,0 % à domicile, 13,5 % en maison de retraite et 1,3 % sur la voie publique.

Tableau 5- Résultats des estimations du taux de surmortalité

Modèle	MCO	SAR	SEM
Densité de population	0.105*** (0.0178)	0.0689*** (0.0237)	0.0893*** (0.0275)
Part d'ouvriers	4.863*** (1.800)	3.652*** (0.943)	4.107*** (1.129)
Part d'inactifs	-0.0707 (1.006)	-0.575 (0.855)	-1.445 (1.192)
Ecart de niveau de vie	0.128 (0.103)	0.110** (0.0483)	0.103* (0.0527)
Nombre de services d'urgence	-0.0547** (0.0228)	-0.0282 (0.0266)	-0.0151 (0.0291)
Part des résidences secondaires	0.101 (0.141)	0.145 (0.175)	0.116 (0.217)
Département rural	0.0394 (0.0390)	0.0438 (0.0328)	0.0494 (0.0331)
Constant	-1.351*** (0.426)	-0.987*** (0.233)	-1.018*** (0.293)
ρ		0.384*** (0.0886)	
λ			0.458*** (0.119)
Observations	96	96	96
R ²	0.516	0.553	0.492
Log Likelihood	56.932	64.906	62.468
AIC	-97.864	-109.812	-104.936
Tests			
I de Moran	3.580***		
LMError	8.317***		
RLMError	0.964		
LMLag	17.110***		
RLMLag	9.758***		

Ecart-types entre parenthèses

***Significatif à 1%, **Significatif à 5%, * Significatif à 10%

6. Conclusion

Cet article propose une analyse de l'hétérogénéité et de l'autocorrélation spatiale des hospitalisations et décès liés à la Covid-19 intervenus entre les semaines 9 et 14 de l'année 2020 et de la surmortalité au 30 mars de la même année au niveau des départements de France métropolitaine. L'inégale répartition des cas et des décès sur le territoire national et le clustering entre départements rapprochés apparaissent fortement pour ce qui concerne directement la maladie et sont confirmés, bien que dans une moindre mesure, pour la surmortalité. Les estimations économétriques réalisées permettent d'expliquer une partie de la diversité géographique observée grâce à l'intégration de variables démographiques, économiques et de cadre de vie et à la prise en compte des effets de débordement entre départements autorisée par l'intégration aux modèles d'une matrice de distance inverse. Nos résultats montrent que la densité, la part des ouvriers dans la population active et les écarts interdéciles de revenus exercent une influence positive sur les hospitalisations, les décès et la surmortalité mais que la disponibilité de services d'urgences influence uniquement les variables

spécifiques à la Covid-19. Les effets de débordement se révèlent également plus nettement marqués pour les deux indicateurs spécifiques que pour la surmortalité en général.

Cette recherche met en évidence que, au-delà de l'importance des caractéristiques individuelles comme facteurs explicatifs de la probabilité de contracter la Covid-19 et de ses conséquences, les éléments liés au contexte économique, démographique et social interviennent également. Les départements les plus denses, les plus inégalitaires ainsi que ceux dans lesquels la part d'ouvriers est la plus élevée se sont en effet révélés les plus vulnérables. Ces caractéristiques soulignent d'abord la complémentarité entre les politiques de santé d'une part et les politiques sociales et de redistribution de l'autre. Le rôle des services d'urgence comme facteur de réduction des manifestations de l'épidémie va dans le même sens. Il montre en effet qu'en présence d'inégalités les services publics, en l'occurrence de santé, permettent de protéger les populations de la maladie et de réduire les décès. Ensuite, en mettant en évidence le rôle de la densité démographique, cet article rappelle que les départements ruraux ou les moins peuplés ne sont pas les plus vulnérables mais que les territoires métropolitains présentent aussi des faiblesses au niveau de la prise en charge des malades qu'il est important de considérer. Enfin, les disparités locales et les effets de débordement que nous avons mis en évidence vont dans le sens d'une régionalisation poussée de la mise en œuvre des politiques de santé. Menées au plus près des territoires, ces dernières peuvent en effet être en mesure de mieux prendre en considération les effets de proximité géographique et socio-économiques qui prévalent au niveau de leur périmètre d'action et, ainsi, être mieux à même d'affecter les ressources humaines et financières.

Le caractère évolutif des données mobilisées dans cet article et les possibles effets retard de la Covid-19 (augmentation des décès liés à la non-prise en charge d'autres pathologies et/ou diminution des décès liés aux accidents de la route et à la pollution) appellent la réalisation d'autres recherches sur les causes et conséquences socio-économiques de cette épidémie. En particulier, le rapprochement de données individuelles et locales dans le cadre de modèles multiniveaux pourrait permettre de mieux distinguer la part du nombre de malades et de décès relevant de causes propres aux individus ou à leur environnement. Un autre complément pourrait être envisagé grâce à l'utilisation de techniques de *data mining* telles que les *Random forests* permettant de traiter simultanément des variables colinéaires. En effet, l'importante corrélation entre les éléments représentatifs du cadre de vie nous a conduit à n'en retenir qu'un petit nombre. Leur prise en compte conjointe permettrait d'affiner les résultats obtenus.

Références

- Anselin L. , Bera A. K., Florax R., Yoon M. J.(1996). Simple diagnostic tests for spatial dependence, *Regional science and urban economics* 26.1, p. 77–104
- Baumont C., Ertur C., Le Gallo J. (2006) Clubs de convergence et effets de débordements géographiques: une analyse spatiale sur données régionales européennes, 1980-1995. *Économie et Prévision*, 173(2): 111–134.
- Bayer, C and M Kuhn (2020), Intergenerational ties and case fatality rates: A cross country analysis, CEPR Discussion Paper no. 14519.
- Bellamy V. (2017) 594 000 personnes décédées en France en 2016, pour un quart d'entre elles à leur domicile, Insee Focus, n° 95. Disponible en ligne : <https://www.insee.fr/fr/statistiques/3134763>
- Belloc M., Buonanno P., Drago F., Galbiati R. et Pinotti P. (2020) , Cross-country correlation analysis for research on COVID-19, Vox-CEPR Policy Portal. Disponible en ligne : <https://voxeu.org/article/cross-country-correlation-analysis-research-covid-19>
- Cambois E. et Jusot F. (2007) Ampleur, tendance et causes des inégalités sociales de santé et de mortalité en Europe : une revue des études comparatives. *Bulletin épidémiologique hebdomadaire*, n° 2-3, 27 janvier.
- Curtis, S. (2004) Health and Inequality, Londres, SAGE Publications.
- DARES (2020) Activité et conditions d'emploi de la main-d'œuvre pendant la crise sanitaire Covid-19- Synthèse des résultats de l'enquête flash-avril 2020. Disponible en ligne : https://dares.travail-emploi.gouv.fr/IMG/pdf/dares_acemo_covid19_synthese_17-04-2020.pdf
- Dietz W., Santos-Burgoa S. (2020), Obesity and its Implications for COVID-19 Mortality, *Obesity*, first published: 01 April 2020 - <https://doi.org/10.1002/oby.22818>
- Farmer P. (1996) Social inequalities and emerging infectious diseases. *Emerging Infectious diseases*, Vol. 2, n°4, pp. 259–269.
- Florax R. J. G. M., Nijkamp P. (2004) Misspecification in Linear Spatial Regression Models ; in Kempf-Leonard K. (Ed.), 2004, *Encyclopedia of Social Measurement*, Academic Press, San Diego.
- Frohlich N. et Mustard C. (1996) A regional comparison of socioeconomic and health indices in a Canadian province, *Social Science & Medicine*, Vol. 42, n°9, pp. 1273-1281
- Geronimus A.T., Bound J. et Waidmann T.A. (1999), Poverty, time, and place: variation in excess mortality across selected US populations, 1980-1990. *Journal of Epidemiology and Community Health*, Vol. 53, n°6, pp. 325-334
- Ghosn W., Kassié D., Jouglu E., Salem G., Rey G., Rican S. (2012), Trends in geographic mortality inequalities and their associations with population changes in France, 1975-2006, *European Journal of Public Health*, vol. 23, n°5, 834-840.
- Goldman N. (2001) Social inequalities in health disentangling the underlying mechanisms; *Annals of the New York Academy of Science*, décembre, pp. 118-139.
- Guillain R, Le Gallo J. (2010) Agglomeration and dispersion of economic activities in Paris and its surroundings. *Environment and Planning B:Urban Analytics and City Science*, 37(6): 961–981.
- Haining R-P (1990) *Spatial data analysis in the social and environmental sciences*. Cambridge University Press.
- Hallepée S., Maurou A. (2019) Quels sont les salariés concernés par le télétravail ?. DARES analyses, n°51, novembre. Disponible en ligne : <http://www.epsilon.insee.fr/jspui/bitstream/1/110047/1/2019-051.pdf>
- Hsueh Y-H., Lee J. et Beltz L. (2012) Spatio-temporal patterns of dengue fever cases in Kaoshiung City, Taiwan, 2003–2008, *Applied Geography*, Vol. 34, mai, pp. 587-594
- Hu Y. et Goldman N. (1990) Mortality Differentials by Marital Status: An International Comparison; *Demography*, Vol. 27, pp. 233–250.
- Karlsdotter K., Martín Martín J., PuertoLópez del Amo González M. (2012) Multilevel analysis of income, income inequalities and health in Spain, *Social Science & Medicine*, Vol. 74, n° 7, pp. 1099-1106

- Le Gallo J. (2002). Économétrie spatiale : l'autocorrélation spatiale dans les modèles de régression linéaire, *Economie & prévision*, 4, p. 139–157.
- Link B.G., Phelan J.C., (1995), Social conditions as fundamental causes of disease, *Journal of Health and Social Behaviour*, Numéro spécial "Forty Years of Medical Sociology: The State of the Art and Directions for the Future", pp. 80-94.
- Mamelund, S-E. (2017) Social inequality – a forgotten factor in pandemic influenza preparedness; *Tidsskr Nor Lægeforen*, Vol. 137, n° 12-13, pp. 911-913
- Olson N.A., Davidow A.L., Winston C.A., Chen M.P., Azmararian J.A. et Katz D.J. (2012) A national study of socioeconomic status and tuberculosis rates by country of birth, United States, 1996–2005, *BMC Public Health*, Vol., Article n°: 365, <https://doi.org/10.1186/1471-2458-12-365>
- Onder G., Resa G., Brusaferro S. (2020), Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy; *JAMA*. Published online March 23, 2020. doi:10.1001/jama.2020.4683
- Prior L., Manley D. et Sabel C.E. (2018), Biosocial health geography: New 'exposomic' geographies of health and place, *Progress in Human Geography*, Vol.43, n°3, pp.531-552.
- Roques L., Klein E., Papaix J. et Soubeyrand S. (2020) Mechanistic-statistical SIR modelling for early estimation of the actual number of cases and mortality rate from COVID-19. *Quantitative Biology - Populations and Evolution*, arXiv:2003.10720v2
- Rosenberg M.W. (1998), Medical or health geography? Populations, peoples and places. *International Journal of Population Geography*, Vol. 4, n°3, pp. 211-226.
- Scott A. et Coote W. (2007) Whither Divisions of General Practice? An empirical and policy analysis of the impact of Divisions within the Australian health care system. *The Medical Journal of Australia*, Vol.187, n°2, pp. 95-99. <https://doi.org/10.5694/j.1326-5377.2007.tb01149.x>
- Scott A., Schurer S., Jensen P.H. et Sivey P. (2009), The effects of an incentive program on quality of care in diabetes management, *Health Economics*, 18, 9, pp. 1091-1108
- Talavera A. et Perez E.M. (2009), Is cholera disease associated with poverty? *Journal of Infection in Developing Countries*, Vol.1-3, n°6, pp. 408-411.
- Tarwater P.M. et Martin C.F. (2001) Effects of population density on the spread of disease, *Complexity*, Vol. 6, n°6, pp. 29-36
- Upton G-J-G., Fingleton B. (1985) *Spatial data analysis by example*. John Wiley & Son, New York.
- Wu Z., McGoogan J.M. (2020), Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China - Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA*. 323(13): 1239-1242. doi:10.1001/jama.2020.2648

Annexe

Statistiques descriptives

Variable	Obs	Mean	Std. Dev.	Min	Max
Taux de décès	96	0,12	0,14	0,00	0,79
Taux d'hospitalisation	96	0,91	0,77	0,13	4,09
Taux de surmortalité	96	0,01	0,19	-0,30	1,08
Densité de population	96	4,65	1,24	2,69	9,94
Taux des ouvriers	96	0,11	0,02	0,04	0,14
Taux des inactifs	96	0,13	0,02	0,09	0,18
Ecart de niveau de vie	96	3,22	0,49	2,60	6,30
Nombre de services d'urgence	96	1,77	0,64	0,00	3,14
Part des résidences secondaires	96	0,11	0,10	0,01	0,45
Département rural	96	0,57	0,50	0,00	1,00

Matrice des corrélations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Densité de population	1.000						
(2) Taux des ouvriers	-0.498***	1.000					
(3) Taux des inactifs	0.521***	-0.392***	1.000				
(4) Ecart niveau de vie	0.777***	-0.643***	0.520***	1.000			
(5) Nombre de services d'urgence	0.634***	-0.307***	0.408***	0.392***	1.000		
(6) Part des résidences secondaires	-0.403***	-0.257**	-0.179*	-0.098	-0.282***	1.000	
(7) Département rural	0.528***	-0.343***	0.588***	0.462***	0.430***	-0.252**	1.000

VIF = 2,5

Covid-19 : analyse spatiale de l'influence des facteurs socio-économiques sur la prévalence et les conséquences de l'épidémie dans les départements français

Mounir AMDAOU

CEPN, CNRS, université Paris Nord
Chercheur rattaché à EconomiX
Mail : mounir.amdaoud@economix.fr

Giuseppe ARCURI

Université Paris 1 Panthéon Sorbonne, PRISM
EconomiX
Mail : giuseppe.arcuri@univ-paris1.fr

Nadine LEVRATTO (auteur correspondant)

EconomiX, CNRS, université Paris Nanterre
Mail : nadine.levratto@parisnanterre.fr

Résumé : Cet article s'intéresse aux déterminants socio-économiques des hospitalisations et des taux de décès liés à la Covid-19 entre les semaines 9 et 15 d'une part et à la surmortalité observée cette année par rapport aux précédentes, d'autre part. Il propose une approche territoriale de ces questions grâce au recours à des données calculées au niveau des départements français, publiées par Santé France et l'INSEE. L'analyse spatiale exploratoire réalisée révèle d'abord l'hétérogénéité et l'autocorrélation spatiale de la maladie et de ses conséquences. Le recours à des modèles d'économétrie spatiale permet ensuite de mettre en évidence l'influence de la densité démographique, des inégalités sociales, de la part des ouvriers dans la population active et des services d'urgence sur les phénomènes étudiés. Enfin, nous rendons également compte du rôle des effets de débordements entre départements situés à proximité les uns des autres. Ces derniers interviennent davantage dans la diffusion des cas et décès liés à la Covid-19 que dans la surmortalité.

Mots clés : Covid-19, déterminants locaux, analyse spatiale

Codes JEL : C21, I14, R23

1. Introduction

L'apparition, la diffusion et le nombre de décès liés à la Covid-19 en France à partir de la fin du mois de février 2020 a suscité de nombreux commentaires tenant d'abord à leur nombre, au taux de croissance et aux classes d'âge les plus concernées. Très rapidement une autre dimension, régionale, cette fois a été soulignée. L'emploi du terme clusters, pour décrire les foyers d'infection, et le focus sur le Grand Est, les Hauts-de-France et l'Île-de-France qui concentrent le plus grand nombre de cas conduisent à se demander de quels mécanismes répond la répartition spatiale de ce virus.

Les facteurs individuels sont évidemment essentiels pour expliquer la probabilité de contracter la maladie et son issue. Cependant, l'importance des facteurs socioéconomiques locaux comme facteur explicatif de l'état de santé des populations et des taux de mortalité est démontrée par la littérature (voir la revue de littérature de Cambois et Jusot, 2007). Elle est également prise en considération par les Etats comme le rappellent les Objectifs du millénaire pour le développement adoptés en 2000. Ces derniers reposent sur l'idée que la lutte contre la pauvreté, à côté de l'amélioration de la médecine et des systèmes de santé, a des conséquences sur la mortalité, maternelle et infantile notamment. La relation entre la situation économique et sociale et la santé est confirmée pour les pays en développement et les pays développés (Link et Phelan, 1995), dont la France (Ghosn et al., 2012).

Cet article s'intéresse à la répartition spatiale du nombre de cas de Covid-19 nécessitant une hospitalisation et du nombre de décès liés à ce virus recensés dans les départements français. Les données utilisées proviennent de l'INSEE et du ministère de la santé. Les premières analyses à ce sujet ont mis en évidence une forte hétérogénéité spatiale dans répartition du nombre d'individus concernés entre les différents départements de France métropolitaine. C'est pourquoi nous proposons de tester économétriquement l'influence des facteurs socioéconomiques locaux susceptibles d'expliquer la prévalence et la mortalité liée à la maladie et d'analyser le poids respectif des caractéristiques démographiques, des niveaux et de la répartition des revenus et des conditions de vie. Dans la mesure où l'épidémie s'est diffusée sur une partie du territoire et que sa propagation a fait l'objet d'une attention considérable de la part des autorités publiques, nous cherchons également à mettre en évidence la structure géographique des données. Pour ce faire, nous recourons à des techniques d'analyse spatiale permettant de repérer les groupes de départements particulièrement touchés ou épargnés.

L'objectif de cette recherche est donc de contribuer à la compréhension des vecteurs de propagation de l'épidémie et d'en souligner la dimension socio-économique. Elle s'inscrit dans le champ de la géographie de la santé qui a émergé dans les années 1970 et a connu un profond renouvellement dans les années 1990 sous l'influence de travaux de recherche nombreux et variés des domaines de l'épidémiologie, des sciences de l'environnement, des sciences sociales, de la santé publique ou du

management des services dans le but de rendre compte des disparités spatiales en matière d'accès aux soins, d'exposition aux maladies et de morbidité (Rosenberg, 1998 ; Curtis, 2004 ; Prior et al. 2018). Cette recherche complète aussi celles entreprises par les épidémiologistes qui pointent les facteurs individuels tels sur l'âge, l'indice de masse corporelle ou l'état de santé des personnes touchées par la Covid-19 (Onder, Resa et Brusaferro, 2020 ; Wu et McGoogan, 2020 ; Diez et Santos-Burgoa, 2020) comme le recense la base de données Worldometer¹.

La prise en considération d'éléments de contexte dans l'analyse d'une question liée à la santé est nécessaire à plusieurs niveaux. Tout d'abord, parce que les différences d'un territoire à l'autre en matière de santé ou de maladie sont de plus en plus souvent pointées par la littérature et que leur omission conduirait à une compréhension partielle des phénomènes étudiés comme le soulignent Geronimus et al. (1999) ou, plus récemment et à propos de la Covid-19 en Italie, Bayer et Kuhn (2020)². Ces derniers envisagent ainsi que les structures familiales et la présence au sein d'un même logement de familles comptant plusieurs générations diffèrent selon les régions et expliquerait ainsi les écarts géographiques observés. Ensuite, car la responsabilité des différents types de caractéristiques d'un territoire peuvent jouer différemment selon le type d'affection. Ainsi, une importante littérature pointe la responsabilité de la pauvreté dans la prévalence du choléra comme l'ont montré Talavera et Perez (2009) dans une comparaison internationale réalisées à partir de données de la Banque Mondiale ou Olson et al. (2012) pour la tuberculose. La qualité du système de santé peut également expliquer les écarts entre régions (Scott et Coote, 2007) comme c'est par exemple le cas pour le diabète (Scott et al. 2009). Enfin, la démographie intervient comme un facteur de différenciation spatiale en matière de santé et, à ce titre, fait l'objet de fréquentes analyses pour expliquer les écarts internationaux (Hu et Goldman, 1990) et régionaux (Frohlich et Mustard, 1996). La troisième raison qui légitime une approche socio-économique à l'échelle géographique de la Covid-19, est l'organisation du système de santé. A partir de l'entrée en vigueur de la loi relative à la politique de santé publique en 2004, l'échelon local est devenu le niveau pertinent en matière de santé publique. La loi portant réforme de l'hôpital et relative aux patients, à la santé et aux territoires (HPST), du 21 juillet 2009, a créé les Agences régionales de santé (ARS) dans un souci de gestion de la politique de santé au plus près du terrain et a introduit la notion de Contrat local de santé (CLS) afin d'articuler les politiques régionales de santé, l'action des collectivités territoriales et des autres acteurs locaux autour d'un diagnostic partagé. De même, le territoire est présent dans la réorganisation hospitalière avec la création des Groupements hospitaliers de territoires (GHT).

¹ Disponible en ligne : <https://www.worldometers.info/coronavirus/>

² A ce stade, il s'agit de pistes de réflexion qui nécessiteront des analyses approfondies lorsque de nouvelles données seront disponibles. Cette hypothèse et les tests réalisés à son sujet sont discutés par Buonnanno et al. (2020).

Notre démarche empirique repose sur les cartes de diffusion de la Covid-19 par département communiquées par l'INSEE et le Ministère des Solidarités et de la Santé depuis le début du mois de mars 2020 et des commentaires d'experts, analystes et journalistes réalisés à chaud et relayés par la presse. Nous en résumons le contenu. Dès le début, les mouvements des populations aisées des grandes métropoles, notamment Paris, vers leurs résidences secondaires localisées dans les départements de Normandie et de Bretagne et de Vendée, ont été pointés comme facteurs de diffusion de l'épidémie. Plusieurs interventions ont également mis l'accent sur la densité comme facteur explicatif des taux élevés de Coronavirus, en Ile de France notamment, et sur la situation des hôpitaux publics et de la médecine de ville. Enfin, plus récemment, le rôle de la pauvreté a été relevé par la presse pour expliquer le non-respect du confinement, en Seine Saint Denis notamment.

Suivant cette logique, nous mobilisons trois grandes familles de variables locales pouvant potentiellement affecter le taux de malades et le taux de mortalité liés à la Covid-19 : des déterminants économiques, des déterminants démographiques et des déterminants liés au cadre de vie. Afin de rendre compte des effets de diffusion et du rôle des différentes familles de déterminants analysés dans ce processus, nous menons une analyse exploratoire de l'autocorrélation spatiale qui permet de mesurer la dépendance spatiale entre les valeurs d'une même variable en différents endroits de l'espace. En effet, les données par département, comme nombre de variables pour lesquelles on dispose d'informations géolocalisées, se caractérisent par des dépendances spatiales qui sont d'autant plus fortes que les localisations sont plus proches. Tenir compte de cette particularité permet donc de prendre en compte les interactions et les externalités spatiales dans l'analyse des phénomènes observés.

Les résultats obtenus mettent en évidence le rôle des facteurs socio-économiques dans les manifestations de la Covid-19 quel que soit l'indicateur obtenu. La densité, la part d'ouvriers dans la population active et les inégalités de revenus influencent positivement le taux d'hospitalisation, le taux de décès et la surmortalité. Le nombre de services d'urgence corrigé de la taille du département exerce en revanche une influence négative sur les trois variables expliquées retenues. Enfin, aucune influence significative de la part des résidences secondaires n'est mise en évidence, ce qui nous conduit à relativiser l'importance des mouvements de population intervenus au moment du confinement.

Le reste de l'article s'organise de la façon suivante. La section 2 met en évidence l'hétérogénéité et le *clustering* spatial de la Covid-19. La section 3 présente les données et les variables explicatives introduites dans l'analyse. La section 4 expose les techniques d'estimation retenues et la section 5 présente et commente les résultats obtenus. Enfin, la section 6 conclut.

2. Hétérogénéité spatiale des cas de Covid-19 et de la mortalité associée

Cette section rend compte de l'hétérogénéité spatiale de l'épidémie de Covid-19 et des taux de mortalité associés. Elle repose sur des données au niveau des départements récemment mises à disposition par le ministère des Solidarités et de la Santé³ et l'INSEE⁴. Notre analyse couvre les 96 départements de France métropolitaine. Les départements d'outre-mer sont exclus de l'analyse car le nombre de cas recensés restait trop faible au moment où les données ont été collectées, de l'absence de certaines données mobilisées dans cette recherche et de la discontinuité géographique. Pour chaque département, trois indicateurs d'intensité de l'épidémie ont été calculés :

- i) un indicateur classique de prévalence mesuré par le nombre d'hospitalisations liés à la Covid-19. Il est défini par la somme des nouvelles hospitalisations entre le 19 mars et le 13 avril rapporté au nombre d'habitants du département pour 1000 habitants ;
- ii) le taux de mortalité mesuré par le nombre de décès dus à la Covid-19 à l'hôpital⁵ recensés le 12 avril rapporté au nombre d'habitants du département pour 1000 habitants ;
- iii) un taux de surmortalité mesuré par la différence entre le nombre total de décès par département au 30 mars 2020⁶ et la moyenne des décès recensés à la même date en 2018 et 2019 rapportée à cette dernière qui permet de tenir compte des décès liés au virus en dehors du milieu hospitalier, dans les EHPAD notamment⁷.

Cette dernière variable a été construite par référence à la moyenne des deux années qui précèdent en raison de la durée exceptionnellement longue de l'épisode de grippe en 2018 et de sa relative faiblesse en 2019. Ainsi au niveau national, le nombre de décès totaux survenus entre le 1er et le 30 mars 2020 est supérieur à celui enregistré sur la même période en 2019 : 57 441 décès ont été enregistrés en 2020 en France contre 52 011 en 2019. Ce nombre reste néanmoins inférieur au nombre des décès (58 641) enregistrés sur la même période en 2018. Nous n'avons pas retenu le taux de létalité mesuré par le nombre de décès rapporté au nombre de cas déclarés en raison des débats portant sur la fiabilité de ce ratio en l'absence de tests à grande échelle. Au 13 avril 2020, on dénombrait 70 265 passages à l'hôpital enregistrés depuis le 1^{er} mars sur la France entière alors que les données actualisées sur les décès en milieu hospitalier en reportent 9 234 au 12 avril.

Les cartes reportées sur les figures 1, 2 et 3 rendent compte de l'hétérogénéité spatiale, très tôt décelée par le secteur médical, de la maladie. Les départements les plus touchés, quel que soit l'indicateur retenu, sont ceux de la région Grand Est, des Hauts-de-France et de l'Ile-de-France. On note

³ Les fichiers sont regroupés sur le site data.gouv.fr.

⁴ Les données sont disponibles à l'adresse suivante : <https://www.insee.fr/fr/information/4470857>

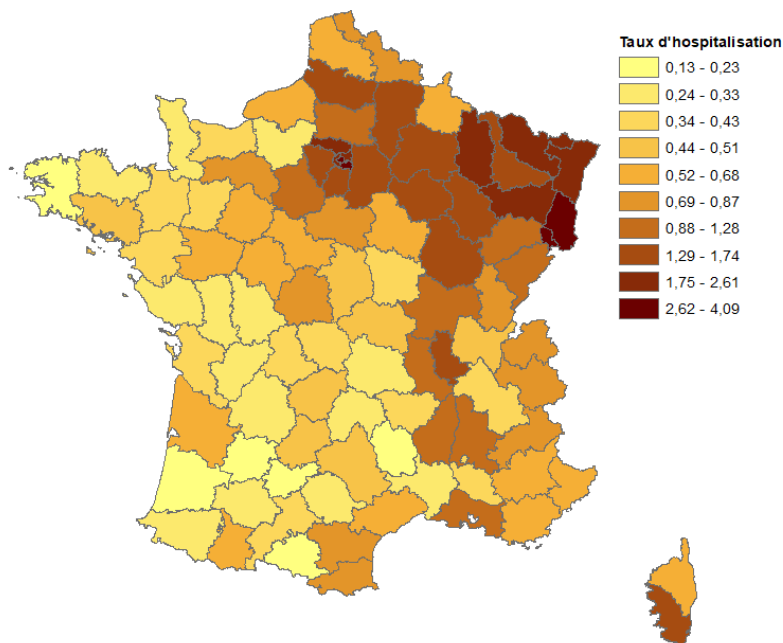
⁵ Le périmètre hospitalier pour le nombre de décès est imposé par la comptabilisation des décès en EHPAD à partir du mois d'avril seulement.

⁶ Les données ont été enregistrées à la date du 10 avril 2020.

⁷ Les données sur ces derniers ont commencées à être intégrées aux statistiques sur les décès à partir du début du mois d'avril.

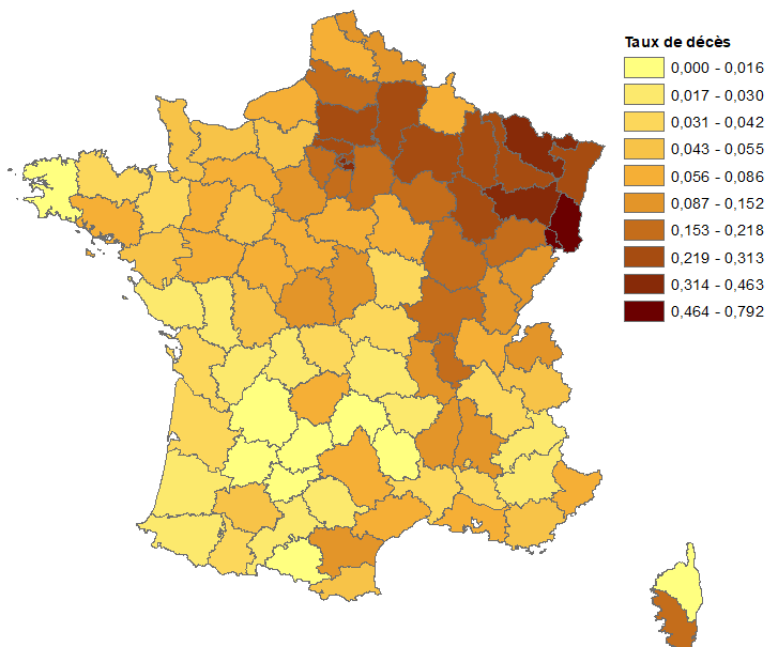
également de plus forts taux d'hospitalisation et de mortalité le long de la vallée du Rhône. En revanche, la moitié ouest du pays apparaît relativement moins touchée par la Covid-19, cette moindre exposition étant particulièrement visible dans le quart sud-ouest.

Figure 1 : Taux d'hospitalisation pour la Covid-19 par département (par décile)



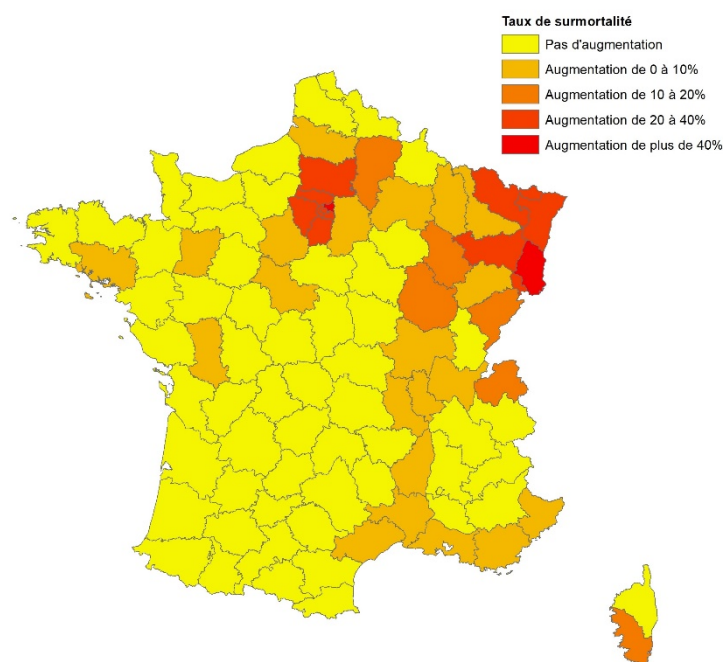
Source : Ministère des solidarités et de la santé. Calculs et cartographie des auteurs

Figure 2 : Taux de décès dus à la Covid-19 en milieu hospitalier par département (par décile)



Source : Ministère des solidarités et de la santé. Calculs et cartographie des auteurs

Figure 3 : Taux de surmortalité par département au 30 mars 2020
par rapport à la moyenne de 2018 et 2019



Source : INSEE. Calculs et cartographie des auteurs

La polarisation dans l'espace des indicateurs analysés peut résulter d'un effet de contagion propageant la maladie d'un département à l'autre. En présence de ce type de regroupement spatial des données, l'utilisation des méthodes habituelles d'analyse s'accompagne d'un risque de biais. En effet, si le phénomène observé dans le département i est influencé par ce qui se passe dans le département j , la condition de normalité des résidus n'est plus respectée. Afin de tester l'existence d'un phénomène de clustering spatial des données, nous appliquons la méthode ESDA (Exploratory Spatial Data Analysis), préconisée pour analyser les données locales (Baumont et al., 2006; Guillain et Le Gallo, 2010).

La détermination des interactions entre les différents départements repose sur la définition d'une matrice de poids spatial (Florax et Nijkamp, 2004). La matrice spatiale W ici utilisée est celle dite de distance inverse⁸. La pondération inverse à la distance est une méthode d'interpolation spatiale classique établie selon un processus permettant d'assigner une valeur à tout point d'un espace à partir d'un semis de points connus, ici les centroïdes des départements.

L'autocorrélation spatiale est mesurée par la statistique I de Moran (Upton et Fingleton, 1985 ; Haining, 1990). Elle représente pour chaque variable x le degré d'association linéaire entre sa valeur à

⁸ Les matrices de distance inverse prennent en compte la relation décroissante entre l'intensité de la relation et la distance. Elles prennent la valeur 1 si $d < d_0$, $1/d^2$, e^{-2d} , etc.

une certaine localisation et la moyenne pondérée spatialement de ses voisins. Cette statistique s'écrit de la façon suivante :

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} (y_{it} - \bar{y}_t) * (y_{jt} - \bar{y}_t)}{\sum_i (y_{it} - \bar{y}_t)^2}$$

dans laquelle I est l'indice d'autocorrélation spatiale de Moran, w_{ij} , est un élément de la matrice de poids spatial W . Les valeurs varient entre -1 (dispersion parfaite) à +1 (corrélation parfaite). Une valeur nulle est significative d'un modèle spatial parfaitement aléatoire. Les valeurs négatives (respectivement positives) de l'indice indiquent une autocorrélation spatiale négative (respectivement positive).

Le Tableau 1 présente les valeurs de la statistique I de Moran des trois indicateurs d'intérêt dans les 96 départements de France métropolitaine aux dates retenues pour cette recherche.

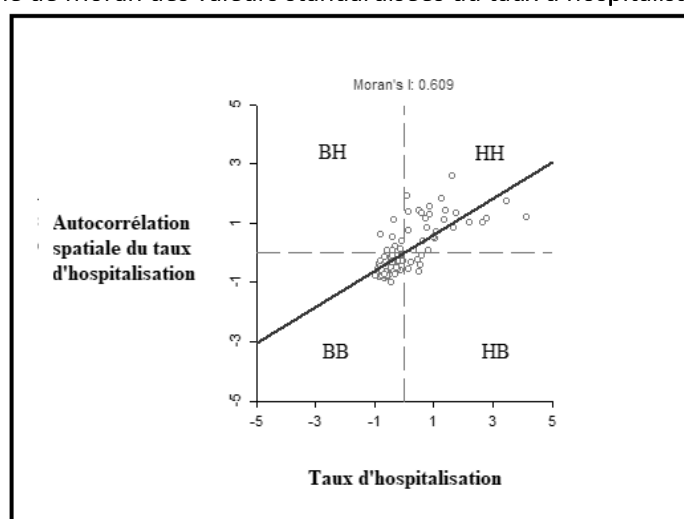
Tableau 1 : Statistique I de Moran

Indicateur	I de Moran	Moyenne	Ecart-type	Valeur standardisée	P-value
Taux d'hospitalisation	0.609	-0.0103	0.0756	8.1968	0.0001
Taux de décès en milieu hospitalier	0.642	-0.0102	0.0738	8.8402	0.0001
Taux de surmortalité	0.560	-0.0112	0.0731	7.7982	0.0001

Source : Santé Publique France et INSEE. Calculs des auteurs.

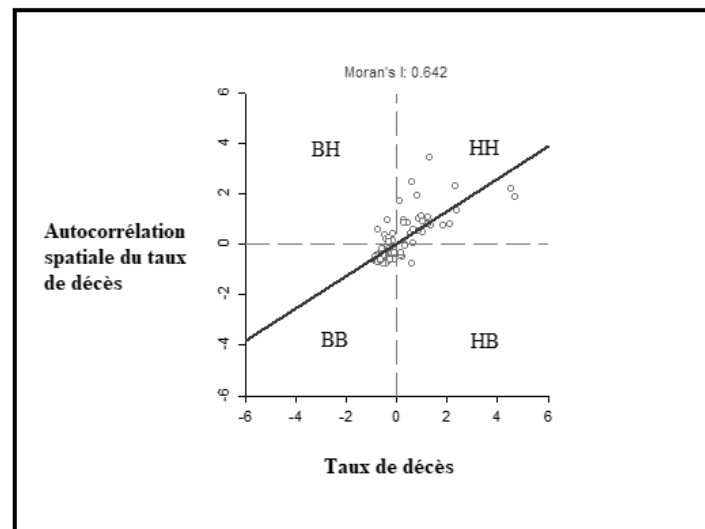
Les valeurs des coefficients obtenues 0.609, 0.642 et 0.560, indiquent la présence d'une importante autocorrélation spatiale positive. Ainsi, les départements qui présentent des taux d'hospitalisation, de mortalité et de surmortalité élevés (respectivement faibles) ont tendance à se trouver à proximité de départements présentant des valeurs des taux similaires. Les diagrammes des figures 4, 5 et 6 rendent compte de ces regroupements spatiaux.

Figure 4 : Diagramme de Moran des valeurs standardisées du taux d'hospitalisation lié à la Covid-19



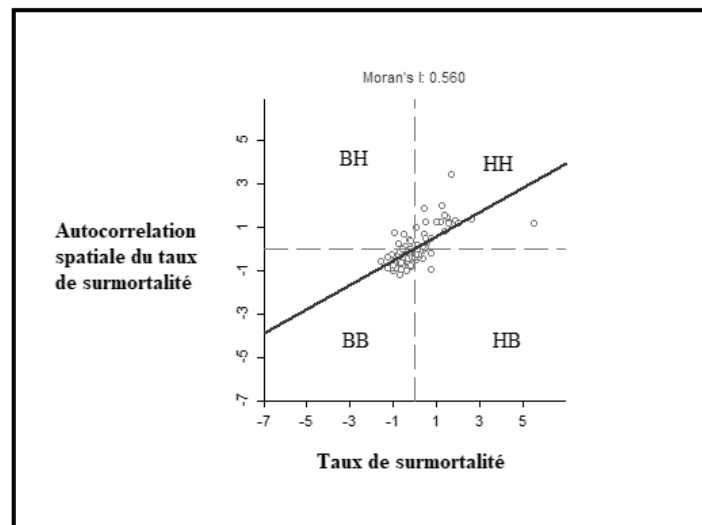
Source : Santé Publique France. Calculs des auteurs

Figure 5 : Diagramme de Moran des valeurs standardisées du taux de décès en milieu hospitalier lié à la Covid-19



Source : Santé Publique France. Calculs des auteurs

Figure 6 : Diagramme de Moran des valeurs standardisées du taux de surmortalité au 30 mars 2020 par rapport à la moyenne de 2018 et 2019



Source : INSEE. Calculs des auteurs

Les LISA (Local Indicator of Spatial Association) proposés par Anselin (1995) viennent compléter l'indice I de Moran en offrant une vision plus précise du *clustering* local. Ils permettent de tester l'hypothèse de distribution aléatoire en comparant les valeurs de chaque localisation spécifique avec les valeurs dans les localisations voisines.

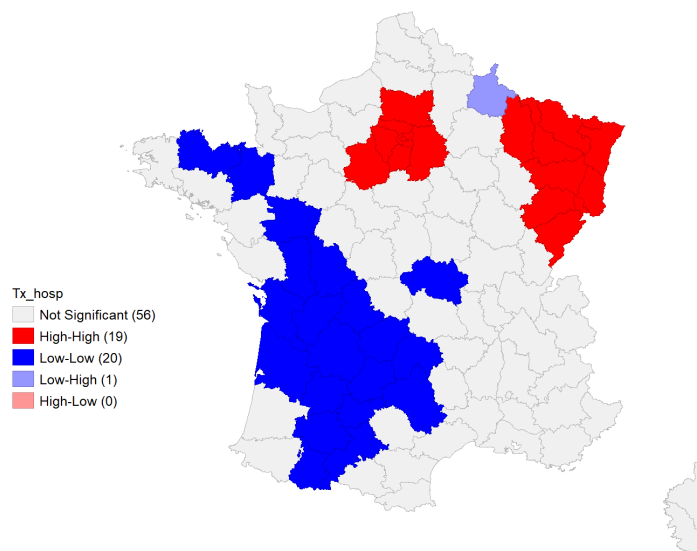
Cinq possibilités sont identifiées :

- Les regroupements de type haut-haut (HH), ou points chauds, où l'autocorrélation spatiale est positive tels que les unités spatiales présentant une valeur élevée de la variable sont entourées d'unités spatiales similaires,

- Les regroupements de type bas-bas (BB), ou points froids, où l'autocorrélation spatiale est également positive mais, qui à la différence des précédents sont tels que les unités spatiales présentant une valeur faible de la variable sont entourées d'unités spatiales similaires,
- Les regroupements atypiques de type haut-bas (HB), où l'autocorrélation spatiale est négative tels que les unités spatiales présentant une valeur élevée de la variable sont entourées d'unités spatiales dans lesquelles la variable prend une valeur faible
- Les regroupements atypiques de type bas-haut (BH), où l'autocorrélation spatiale est également négative mais que sont tels que les unités spatiales présentant une valeur faible de la variable sont entourées d'unités spatiales dans lesquelles la variable prend une valeur forte,
- Les unités spatiales pour lesquelles aucune autocorrélation spatiale significative n'est repérée.

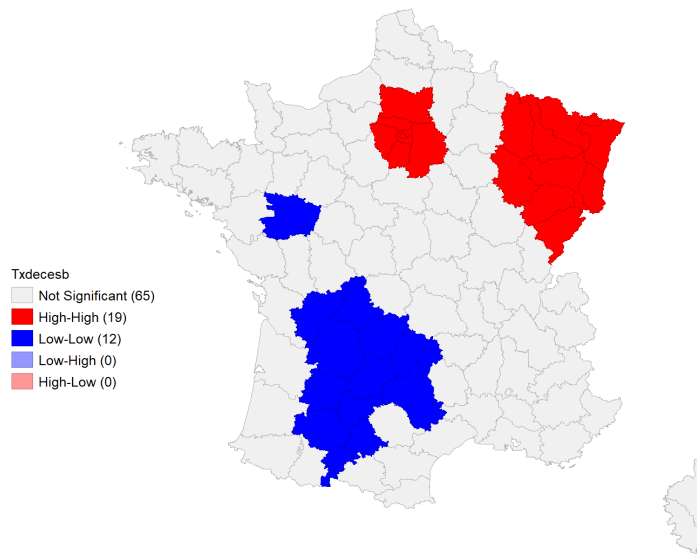
Les cartes des figures 7, 8 et 9 représentent les autocorrélations spatiales des taux d'hospitalisation, de mortalité et de surmortalité par département données par les statistiques LISA.

Figure 7 : LISA-Taux d'hospitalisation de la Covid-19 par département



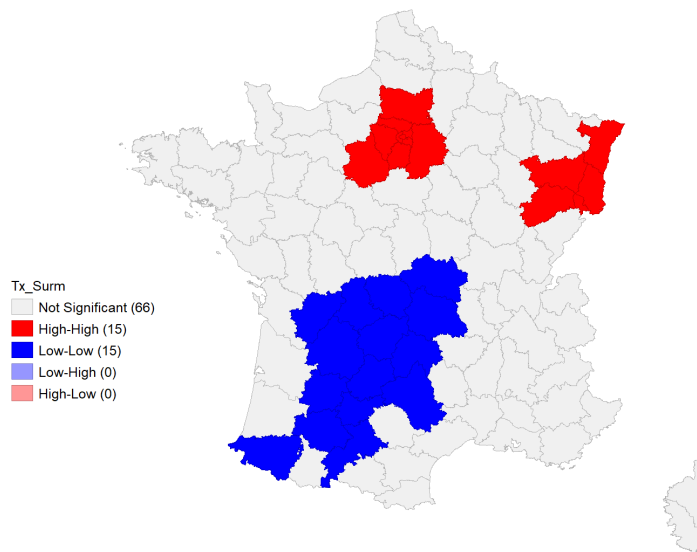
Source : Santé Publique France. Calculs des auteurs.

Figure 8 : LISA-Taux de décès liés à la Covid-19 par département



Source : Santé Publique France. Calculs des auteurs.

Figure 9 : LISA-Taux de surmortalité par département



Source : INSEE. Calculs des auteurs.

La carte des LISA pour les taux d'hospitalisation rend compte de la présence de deux clusters de type HH situés dans l'est de la France et dans le bassin parisien. Un grand quart sud-ouest et une partie de la Bretagne sont au contraire caractérisés par des taux d'hospitalisation significativement plus faibles que sur le reste du territoire. Ces départements forment un grand cluster de type BB. A l'exception du département des Ardennes caractérisé par une autocorrélation négative rapportée au cluster du nord-est, aucune autocorrélation spatiale n'est décelée sur le reste du pays. La carte des taux de mortalité présente une structure similaire pour le cluster de type HH mais le cluster de type BB est de moindre

ampleur. Il touche seulement le sud-ouest tandis que la Bretagne et le Centre deviennent non significatifs. Cette même structure d'ensemble est observée pour les LISA de la surmortalité mais on note que les départements composant le cluster HH à l'est sont, dans ce cas, moins nombreux.

3. Identification des variables explicatives

Notre base de données est constituée des 96 départements de France métropolitaine. Dans les analyses qui suivent, nous utilisons soit des ratios pondérés par la population pour tenir compte des différences de taille entre les départements⁹, soit des distributions pour refléter les disparités pouvant exister au sein d'un même département.

Nous distinguons trois natures de variables définies à l'échelle du département susceptibles d'expliquer l'hétérogénéité spatiale des cas et des décès liés à la Covid-19 : des déterminants démographiques, des déterminants économiques et des déterminants liés au cadre de vie. La plupart des informations sont fournies par l'INSEE. Nous avons considéré la dernière année disponible au moment de la rédaction de cet article qui, selon les cas est 2016 ou 2018.

Afin de comprendre l'intensité de la Covid-19 et ses conséquences sur la mortalité au sein de chaque département d'une part et sa diffusion spatiale de l'autre, nous intégrons à notre modèle les variables démographiques suivantes : la densité de population au km², la part des ouvriers dans la population totale et le taux d'inactifs. La densité est prise en compte en tant que facteur aggravant de la contagion (Tarwater et Martin, 2001), la part des ouvriers plus exposée à la continuité du travail (DARES, 2020) et donc à la maladie ainsi que le taux d'inactifs qui capte les personnes âgées, plus affectées par la maladie, et les étudiants scolarisés, considérés comme des vecteurs de propagation de la Covid-19.

Nous les complétons par une variable économique qui est l'écart interdécile de la distribution des revenus. Elle est introduite en tant que proxy des inégalités économiques qui caractérisent les départements et exercent une influence sur l'émergence et la propagation des maladies (Farmer, 1996, Mamelund, 2017) en raison des différences d'accès aux aménités selon le type de population. Trois variables sur le cadre de vie sont considérées. D'abord, la part des résidences secondaires introduite en raison des débats publics sur le rôle des mouvements de population des grandes métropoles, notamment, vers les départements plus touristiques suite à l'annonce du confinement. Ensuite le nombre de services d'urgences par département pour rendre compte de la capacité d'accueil des malades et une variable binaire distinguant les départements urbains et ruraux construite à partir du classement fourni par l'Assemblée des départements de France (ADF).

⁹ En 2016, on dénombrait 76 422 habitants en Lozère et 2 603 723 dans le Nord

Le tableau 2 ci-dessous résume la définition des différentes variables intégrées dans nos modèles. Les statistiques descriptives et la matrice des corrélations se trouvent en annexe.

Tableau 2 : Définition, année de référence et source des variables

Libellé	Définition	Date/Année	Source
Variables expliqués			
Taux d'hospitalisation	(Nombre d'hospitalisation Covid-19/population) *1000	19 Mars -12 Avril 2020	Santé Publique France
Taux de décès	(Nombre de décès Covid19 /population) *1000	19 Mars -13 Avril 2020	Santé Publique France
Taux de surmortalité	Taux d'évolution des décès cumulés au 30 mars 2020 par rapport à la moyenne des décès cumulés entre 2018-2019 à la même date.		INSEE
Variables explicatives			
Densité de population	Log (Nombre d'habitants/ km ²)	2016	INSEE
Taux des ouvriers	Part de la population de 15 ans ou plus selon la CSP ¹⁰ Ouvrier (2016)	2016	INSEE
Taux des inactifs	Part de la population de 15 ans ou plus selon la CSP Autres + personnes sans activité professionnelle - (2016)	2016	INSEE
Ecart niveau de vie	Rapport interdécile des revenus (9 ^e décile/1 ^{er} décile)	2016	INSEE
Nombre de services d'urgence	Log (Nombre Service d'urgences)	2018	INSEE
Part des résidences secondaires	Part des résidences secondaires (y compris les logements occasionnels) dans le total des logements	2016	INSEE
Département rural	Variable dichotomique prenant la valeur 0 si le département est rural ou composé de petites villes et 1 sinon.		ADF

4. Modèles mobilisés

La spécification économétrique considérée dans cette recherche prend comme point de départ le modèle de régression linéaire des moindres carrés ordinaires (MCO) :

$$Y = X\beta + \varepsilon \quad (1)$$

Y représente la variable dépendante (taux d'hospitalisation, taux de décès, taux de surmortalité). X représente les variables explicatives utilisées, β est le vecteur de paramètres à estimer et ε est le terme d'erreur. Lorsqu'un phénomène d'autocorrélation spatiale est ignoré dans la spécification du modèle, mais présent dans le processus générateur des données, les estimateurs des MCO sont biaisés et non convergents.

¹⁰ Catégorie socio-professionnelle

Le modèle spatial autorégressif (SAR) consiste à corriger ce biais en intégrant une “variable endogène décalée” WY dans le modèle (1) et à prendre en compte l'autocorrélation spatiale relative à la variable Y . Le modèle s'écrit comme suit :

$$Y = \rho WY + X\beta + \varepsilon \quad (2)$$

WY est la variable endogène décalée pour la matrice de distance inverse W , ρ est le paramètre autorégressif indiquant l'intensité de l'interaction existante entre les observations de Y . Dans ce modèle, l'observation de Y est en partie expliquée par les valeurs prises par Y dans les départements voisins.

L'introduction de la variable WY dans le modèle (1) est un moyen d'apprécier le degré de dépendance spatiale alors que les autres variables sont contrôlées. Symétriquement, il permet de contrôler la dépendance spatiale pour évaluer l'impact des autres variables explicatives.

Une seconde façon d'incorporer l'autocorrélation spatiale dans les modèles économétriques est le Spatial Error Model (SEM) qui consiste à spécifier un processus de dépendance spatiale des erreurs dans un modèle de régression. Le modèle SEM est défini de la façon suivante :

$$Y = X\beta + \varepsilon \text{ avec } \varepsilon = \lambda W\varepsilon + u \quad (3)$$

Le paramètre λ reflète l'intensité de l'interdépendance entre les résidus de la régression et u est le terme d'erreur. Omettre une autocorrélation spatiale des erreurs produit des estimateurs non biaisés mais inefficients, si bien que l'inférence statistique basée sur les MCO sera biaisée.

Suivant Le Gallo (2002), différentes approches peuvent être utilisées pour le choix des modèles. Nous avons retenu l'approche dite ascendante qui consiste à commencer avec le modèle non spatial. Des tests du multiplicateur de Lagrange (Anselin et al., 1996) permettent ensuite de trancher entre le modèle SAR, SEM ou le modèle non spatial¹¹.

5. Résultats : présentation et commentaires

Les résultats des estimations des déterminants du taux d'hospitalisation, de décès et de surmortalité sont présentés dans les tableaux 3, 4 et 5 en tenant compte des différentes spécifications du modèle présentées (MCO, SAR et SEM). Les tests du multiplicateur de Lagrange ici réalisés montrent une meilleure performance du modèle SAR. Nous commentons donc ce dernier.

¹¹ Le choix entre les trois modèles s'effectue en comparant les niveaux de significativité des tests du multiplicateur de Lagrange (LMError, LMLag) et leurs versions robustes (RLMError et RLMLag). Si LMLag est aussi significatif que LMError et que seul RLMLag est significatif, on inclut une variable endogène décalée et le modèle privilégié sera le SAR. Inversement, si LMError est aussi significatif que LMLag et que seul RLMError est significatif, alors on choisit le modèle avec autocorrélation des erreurs (SEM) (Le Gallo, 2002). Si aucun test n'est significatif le modèle MCO doit être retenu.

Les résultats de l'estimation des déterminants du taux d'hospitalisation sont présentés dans le tableau 3. Ils mettent en évidence le rôle positif de la densité, conforme à celui attendu. L'importance de la densité de population dans la Covid-19 a été fortement discutée dès le début de l'épidémie en Chine. Le rôle de cette variable est d'autant plus important qu'il est lié au nombre de reproduction de base (connu sous l'abréviation R_0) souvent estimé entre 2 et 4 (Roques et al., 2020), reconnu comme très élevé par les épidémiologistes. Le taux d'hospitalisation est également positivement et significativement corrélé à la part des ouvriers dans la population active. Cela confirme le risque d'exposition accru de cette catégorie sociale (conducteurs des systèmes de transport, livreurs, services de voirie et de nettoyage dans la fonction publique territoriale, réparations d'urgence dans les industries de réseaux, etc.) en raison des moindres possibilités de pratique du télétravail et de leur participation plus élevée aux plans de continuité d'activité mis en œuvre dès le début du confinement. Les inégalités sociales approximées par l'écart interdécile des revenus influencent également de manière positive et significative le taux d'hospitalisation. On retrouve ici les résultats de publications antérieures qui montrent que les inégalités sociales sont des déterminants de la santé des populations (Karlsdotter et al. 2012 ; Goldman, 2001) et des facteurs de propagation des épidémies (Tarwater et Martin, 2001 ; Hsueh et al. 2012). Enfin, le nombre de services d'urgences exerce une influence significative négative sur le taux d'hospitalisation, semblant ainsi confirmer qu'une bonne prise en charge médicale permet de réduire le risque d'évolution dangereuse de la maladie¹².

Il est à noter que la part des résidences secondaires n'est pas significative. Elles avaient été soupçonnées d'avoir aggravé la propagation de l'épidémie au cours des jours ayant suivi l'annonce du confinement en attirant des ménages en provenance de la région parisienne, très touchée par l'épidémie.

L'autocorrélation spatiale, identifiée dans l'ESDA, se retrouve dans les estimations. La valeur du coefficient ρ est de 0,505 et significative à 1%. Ce résultat traduit un effet de diffusion entre les départements localisés à proximité les uns des autres¹³.

¹² Il n'a pas été possible d'introduire le nombre de médecins pour 1000 habitants en raison de la forte corrélation (0,87) entre cette variable et le nombre de services d'urgences.

¹³ Les résultats obtenus sont robustes au type de matrice spatiale utilisés. Nous les avons calculés avec une matrice de contiguïté d'ordre 1 de type reine. Nous les tenons à la disposition des lecteur sur simple demande.

Tableau 3 : Résultats des estimations du taux d'hospitalisation

Modèle	MCO	SAR	SEM
Densité de population	0.310** (0.130)	0.186** (0.0781)	0.311*** (0.0987)
Part d'ouvriers	13.16** (5.114)	5.683* (3.285)	-0.0368 (4.251)
Part d'inactifs	7.293* (3.910)	3.155 (2.967)	-0.426 (4.242)
Ecart de niveau de vie	0.722** (0.283)	0.439*** (0.170)	0.250 (0.184)
Nombre de services d'urgence	-0.469** (0.200)	-0.322*** (0.0915)	-0.335*** (0.0911)
Part des résidences secondaires	-0.914 (0.759)	-0.325 (0.604)	-0.121 (0.796)
Département rural	-0.0379 (0.160)	0.00759 (0.113)	-0.0127 (0.107)
Constant	-4.244*** (1.174)	-2.259*** (0.802)	-0.697 (1.080)
ρ		0.505*** (0.0760)	
λ			0.675*** (0.0801)
Observations	96	96	96
R ²	0.564	0.608	0.442
Log Likelihood	-71.323	-55.081	-57.187
AIC	158.646	130.162	134.374
Tests			
I de Moran	4.748***		
LMError	15.962***		
RLMError	0.491		
LMLag	31.309***		
RLMLag	15.839***		

Ecart-types entre parenthèses

***Significatif à 1%, **Significatif à 5%, * Significatif à 10%

Les résultats des estimations du taux de décès liés à la Covid-19 en milieu hospitalier (Tableau 4) vont dans le même sens que les précédents. Ils mettent en évidence l'influence positive significative de la densité de population, de la part d'ouvriers dans la population active et des écarts de niveau de vie. Ils révèlent aussi l'influence significative négative du nombre de services d'urgences. Comme précédemment, nos résultats ne montrent pas d'influence de la part d'inactifs, plus souvent confinés que les actifs, du fait de la fermeture des écoles et des règles de prudence renforcée appliquées aux personnes âgées. L'absence de significativité de la part des résidences secondaires est de nouveau observée.

Les taux de décès sont également autocorrélés spatialement comme l'indique la valeur de 0,550 prise par le coefficient ρ .

Tableau 4 : Résultats des estimations du taux de décès en milieu hospitalier

Modèle	MCO	SAR	SEM
Densité de population	0.0532** (0.0259)	0.0324** (0.0149)	0.0544*** (0.0191)
Part d'ouvriers	3.342*** (1.170)	1.631** (0.638)	1.230 (0.797)
Part d'inactifs	1.453* (0.830)	0.470 (0.574)	-0.188 (0.823)
Ecart de niveau de vie	0.129** (0.0626)	0.0764** (0.0325)	0.0590* (0.0351)
Nombre de services d'urgence	-0.0838* (0.0445)	-0.0532*** (0.0176)	-0.0537*** (0.0177)
Part des résidences secondaires	-0.182 (0.115)	-0.0478 (0.116)	-0.0666 (0.153)
Département rural	-0.0196 (0.0334)	-0.00959 (0.0217)	-0.0110 (0.0209)
Constant	-0.905*** (0.270)	-0.477*** (0.155)	-0.329 (0.203)
ρ		0.544*** (0.0738)	
λ			0.662*** (0.0770)
Observations	96	96	96
R ²	0.499	0.550	0.418
Log Likelihood	83.526	102.422	100.523
AIC	-151.051	-184.844	-181.047
Tests			
I de Moran	5.791***		
LMError	24.867***		
RLMError	0.111		
LMlag	38.431***		
RLMLag	13.676***		

Ecart-types entre parenthèses

***Significatif à 1%, **Significatif à 5%, * Significatif à 10%

L'estimation des taux de surmortalité (tableau 5), qui ne sont pas exclusivement liés à l'épidémie, permet cependant de capter les décès à les EHPAD et ceux des personnes non testées mais décédées de la Covid-19. Les résultats obtenus sont similaires aux précédents pour les variables démographiques et la variable économique. L'influence négative des services d'urgence disparaît dès que l'ensemble des décès (environ 600 000 par an selon l'INED) est pris en considération. Leur répartition peut expliquer cette absence de résultats. En effet, en 2016, 59,2 % des décès ont eu lieu dans des établissements de santé (hôpital ou clinique)¹⁴ (Bellamy, 2017) si bien que pour nombre d'entre eux, les services d'urgences ne sont pas un point de passage contrairement à ce qui se produit pour la Covid-19.

La valeur prise par le coefficient ρ est égale à 0.384. L'effet de débordement, observé pour les variables expliquées précédente se retrouve de nouveau ici, mais à un degré moindre.

¹⁴ Le reste se répartit entre 26,0 % à domicile, 13,5 % en maison de retraite et 1,3 % sur la voie publique.

Tableau 5- Résultats des estimations du taux de surmortalité

Modèle	MCO	SAR	SEM
Densité de population	0.105*** (0.0178)	0.0689*** (0.0237)	0.0893*** (0.0275)
Part d'ouvriers	4.863*** (1.800)	3.652*** (0.943)	4.107*** (1.129)
Part d'inactifs	-0.0707 (1.006)	-0.575 (0.855)	-1.445 (1.192)
Ecart de niveau de vie	0.128 (0.103)	0.110** (0.0483)	0.103* (0.0527)
Nombre de services d'urgence	-0.0547** (0.0228)	-0.0282 (0.0266)	-0.0151 (0.0291)
Part des résidences secondaires	0.101 (0.141)	0.145 (0.175)	0.116 (0.217)
Département rural	0.0394 (0.0390)	0.0438 (0.0328)	0.0494 (0.0331)
Constant	-1.351*** (0.426)	-0.987*** (0.233)	-1.018*** (0.293)
ρ		0.384*** (0.0886)	
λ			0.458*** (0.119)
Observations	96	96	96
R ²	0.516	0.553	0.492
Log Likelihood	56.932	64.906	62.468
AIC	-97.864	-109.812	-104.936
Tests			
I de Moran	3.580***		
LMError	8.317***		
RLMError	0.964		
LMlag	17.110***		
RLMLag	9.758***		

Ecart-types entre parenthèses

***Significatif à 1%, **Significatif à 5%, * Significatif à 10%

6. Conclusion

Cet article propose une analyse de l'hétérogénéité et de l'autocorrélation spatiale des hospitalisations et décès liés à la Covid-19 intervenus entre les semaines 9 et 14 de l'année 2020 et de la surmortalité au 30 mars de la même année au niveau des départements de France métropolitaine. L'inégale répartition des cas et des décès sur le territoire national et le clustering entre départements rapprochés apparaissent fortement pour ce qui concerne directement la maladie et sont confirmés, bien que dans une moindre mesure, pour la surmortalité. Les estimations économétriques réalisées permettent d'expliquer une partie de la diversité géographique observée grâce à l'intégration de variables démographiques, économiques et de cadre de vie et à la prise en compte des effets de débordement entre départements autorisée par l'intégration aux modèles d'une matrice de distance inverse. Nos résultats montrent que la densité, la part des ouvriers dans la population active et les écarts interdéciles de revenus exercent une influence positive sur les hospitalisations, les décès et la surmortalité mais que la disponibilité de services d'urgences influence uniquement les variables

spécifiques à la Covid-19. Les effets de débordement se révèlent également plus nettement marqués pour les deux indicateurs spécifiques que pour la surmortalité en général.

Cette recherche met en évidence que, au-delà de l'importance des caractéristiques individuelles comme facteurs explicatifs de la probabilité de contracter la Covid-19 et de ses conséquences, les éléments liés au contexte économique, démographique et social interviennent également. Les départements les plus denses, les plus inégalitaires ainsi que ceux dans lesquels la part d'ouvriers est la plus élevée se sont en effet révélés les plus vulnérables. Ces caractéristiques soulignent d'abord la complémentarité entre les politiques de santé d'une part et les politiques sociales et de redistribution de l'autre. Le rôle des services d'urgence comme facteur de réduction des manifestations de l'épidémie va dans le même sens. Il montre en effet qu'en présence d'inégalités les services publics, en l'occurrence de santé, permettent de protéger les populations de la maladie et de réduire les décès. Ensuite, en mettant en évidence le rôle de la densité démographique, cet article rappelle que les départements ruraux ou les moins peuplés ne sont pas les plus vulnérables mais que les territoires métropolitains présentent aussi des faiblesses au niveau de la prise en charge des malades qu'il est important de considérer. Enfin, les disparités locales et les effets de débordement que nous avons mis en évidence vont dans le sens d'une régionalisation poussée de la mise en œuvre des politiques de santé. Menées au plus près des territoires, ces dernières peuvent en effet être en mesure de mieux prendre en considération les effets de proximité géographique et socio-économiques qui prévalent au niveau de leur périmètre d'action et, ainsi, être mieux à même d'affecter les ressources humaines et financières.

Le caractère évolutif des données mobilisées dans cet article et les possibles effets retard de la Covid-19 (augmentation des décès liés à la non-prise en charge d'autres pathologies et/ou diminution des décès liés aux accidents de la route et à la pollution) appellent la réalisation d'autres recherches sur les causes et conséquences socio-économiques de cette épidémie. En particulier, le rapprochement de données individuelles et locales dans le cadre de modèles multiniveaux pourrait permettre de mieux distinguer la part du nombre de malades et de décès relevant de causes propres aux individus ou à leur environnement. Un autre complément pourrait être envisagé grâce à l'utilisation de techniques de *data mining* telles que les *Random forests* permettant de traiter simultanément des variables colinéaires. En effet, l'importante corrélation entre les éléments représentatifs du cadre de vie nous a conduit à n'en retenir qu'un petit nombre. Leur prise en compte conjointe permettrait d'affiner les résultats obtenus.

Références

- Anselin L. , Bera A. K., Florax R., Yoon M. J.(1996). Simple diagnostic tests for spatial dependence, *Regional science and urban economics* 26.1, p. 77–104
- Baumont C., Ertur C., Le Gallo J. (2006) Clubs de convergence et effets de débordements géographiques: une analyse spatiale sur données régionales européennes, 1980-1995. *Économie et Prévision*, 173(2): 111–134.
- Bayer, C and M Kuhn (2020), Intergenerational ties and case fatality rates: A cross country analysis, CEPR Discussion Paper no. 14519.
- Bellamy V. (2017) 594 000 personnes décédées en France en 2016, pour un quart d'entre elles à leur domicile, Insee Focus, n° 95. Disponible en ligne : <https://www.insee.fr/fr/statistiques/3134763>
- Belloc M., Buonanno P., Drago F., Galbiati R. et Pinotti P. (2020) , Cross-country correlation analysis for research on COVID-19, Vox-CEPR Policy Portal. Disponible en ligne : <https://voxeu.org/article/cross-country-correlation-analysis-research-covid-19>
- Cambois E. et Jusot F. (2007) Ampleur, tendance et causes des inégalités sociales de santé et de mortalité en Europe : une revue des études comparatives. *Bulletin épidémiologique hebdomadaire*, n° 2-3, 27 janvier.
- Curtis, S. (2004) *Health and Inequality*, Londres, SAGE Publications.
- DARES (2020) Activité et conditions d'emploi de la main-d'œuvre pendant la crise sanitaire Covid-19- Synthèse des résultats de l'enquête flash-avril 2020. Disponible en ligne : https://dares.travail-emploi.gouv.fr/IMG/pdf/dares_acemo_covid19_synthese_17-04-2020.pdf
- Dietz W., Santos-Burgoa S. (2020), Obesity and its Implications for COVID-19 Mortality, *Obesity*, first published: 01 April 2020 - <https://doi.org/10.1002/oby.22818>
- Farmer P. (1996) Social inequalities and emerging infectious diseases. *Emerging Infectious diseases*, Vol. 2, n°4, pp. 259–269.
- Florax R. J. G. M., Nijkamp P. (2004) Misspecification in Linear Spatial Regression Models ; in Kempf-Leonard K. (Ed.), 2004, *Encyclopedia of Social Measurement*, Academic Press, San Diego.
- Frohlich N. et Mustard C. (1996) A regional comparison of socioeconomic and health indices in a Canadian province, *Social Science & Medicine*, Vol. 42, n°9, pp. 1273-1281
- Geronimus A.T., Bound J. et Waidmann T.A. (1999), Poverty, time, and place: variation in excess mortality across selected US populations, 1980-1990. *Journal of Epidemiology and Community Health*, Vol. 53, n°6, pp. 325-334
- Ghosn W., Kassié D., Jouglu E., Salem G., Rey G., Rican S. (2012), Trends in geographic mortality inequalities and their associations with population changes in France, 1975-2006, *European Journal of Public Health*, vol. 23, n°5, 834-840.
- Goldman N. (2001) Social inequalities in health disentangling the underlying mechanisms; *Annals of the New York Academy of Science*, décembre, pp. 118-139.
- Guillain R, Le Gallo J. (2010) Agglomeration and dispersion of economic activities in Paris and its surroundings. *Environment and Planning B:Urban Analytics and City Science*, 37(6): 961–981.
- Haining R-P (1990) *Spatial data analysis in the social and environmental sciences*. Cambridge University Press.
- Hallepée S., Maurou A. (2019) Quels sont les salariés concernés par le télétravail ?. DARES analyses, n°51, novembre. Disponible en ligne : <http://www.epsilon.insee.fr/jspui/bitstream/1/110047/1/2019-051.pdf>
- Hsueh Y-H., Lee J. et Beltz L. (2012) Spatio-temporal patterns of dengue fever cases in Kaoshiung City, Taiwan, 2003–2008, *Applied Geography*, Vol. 34, mai, pp. 587-594
- Hu Y. et Goldman N. (1990) Mortality Differentials by Marital Status: An International Comparison; *Demography*, Vol. 27, pp. 233–250.
- Karlsdotter K., Martín Martín J., PuertoLópez del Amo González M. (2012) Multilevel analysis of income, income inequalities and health in Spain, *Social Science & Medicine*, Vol. 74, n° 7, pp. 1099-1106

- Le Gallo J. (2002). Économétrie spatiale : l'autocorrélation spatiale dans les modèles de régression linéaire, *Economie & prévision*, 4, p. 139–157.
- Link B.G., Phelan J.C., (1995), Social conditions as fundamental causes of disease, *Journal of Health and Social Behaviour*, Numéro spécial "Forty Years of Medical Sociology: The State of the Art and Directions for the Future", pp. 80-94.
- Mamelund, S-E. (2017) Social inequality – a forgotten factor in pandemic influenza preparedness; *Tidsskr Nor Lægeforen*, Vol. 137, n° 12-13, pp. 911-913
- Olson N.A., Davidow A.L., Winston C.A., Chen M.P., Azmararian J.A. et Katz D.J. (2012) A national study of socioeconomic status and tuberculosis rates by country of birth, United States, 1996–2005, *BMC Public Health*, Vol., Article n°: 365, <https://doi.org/10.1186/1471-2458-12-365>
- Onder G., Resa G., Brusaferro S. (2020), Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy; *JAMA*. Published online March 23, 2020. doi:10.1001/jama.2020.4683
- Prior L., Manley D. et Sabel C.E. (2018), Biosocial health geography: New 'exposomic' geographies of health and place, *Progress in Human Geography*, Vol.43, n°3, pp.531-552.
- Roques L., Klein E., Papaix J. et Soubeyrand S. (2020) Mechanistic-statistical SIR modelling for early estimation of the actual number of cases and mortality rate from COVID-19. *Quantitative Biology - Populations and Evolution*, arXiv:2003.10720v2
- Rosenberg M.W. (1998), Medical or health geography? Populations, peoples and places. *International Journal of Population Geography*, Vol. 4, n°3, pp. 211-226.
- Scott A. et Coote W. (2007) Whither Divisions of General Practice? An empirical and policy analysis of the impact of Divisions within the Australian health care system. *The Medical Journal of Australia*, Vol.187, n°2, pp. 95-99. <https://doi.org/10.5694/j.1326-5377.2007.tb01149.x>
- Scott A., Schurer S., Jensen P.H. et Sivey P. (2009), The effects of an incentive program on quality of care in diabetes management, *Health Economics*, 18, 9, pp. 1091-1108
- Talavera A. et Perez E.M. (2009), Is cholera disease associated with poverty? *Journal of Infection in Developing Countries*, Vol.1-3, n°6, pp. 408-411.
- Tarwater P.M. et Martin C.F. (2001) Effects of population density on the spread of disease, *Complexity*, Vol. 6, n°6, pp. 29-36
- Upton G-J-G., Fingleton B. (1985) *Spatial data analysis by example*. John Wiley & Son, New York.
- Wu Z., McGoogan J.M. (2020), Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China - Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA*. 323(13): 1239-1242. doi:10.1001/jama.2020.2648

Annexe

Statistiques descriptives

Variable	Obs	Mean	Std. Dev.	Min	Max
Taux de décès	96	0,12	0,14	0,00	0,79
Taux d'hospitalisation	96	0,91	0,77	0,13	4,09
Taux de surmortalité	96	0,01	0,19	-0,30	1,08
Densité de population	96	4,65	1,24	2,69	9,94
Taux des ouvriers	96	0,11	0,02	0,04	0,14
Taux des inactifs	96	0,13	0,02	0,09	0,18
Ecart de niveau de vie	96	3,22	0,49	2,60	6,30
Nombre de services d'urgence	96	1,77	0,64	0,00	3,14
Part des résidences secondaires	96	0,11	0,10	0,01	0,45
Département rural	96	0,57	0,50	0,00	1,00

Matrice des corrélations

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
(1) Densité de population	1.000						
(2) Taux des ouvriers	-0.498***	1.000					
(3) Taux des inactifs	0.521***	-0.392***	1.000				
(4) Ecart niveau de vie	0.777***	-0.643***	0.520***	1.000			
(5) Nombre de services d'urgence	0.634***	-0.307***	0.408***	0.392***	1.000		
(6) Part des résidences secondaires	-0.403***	-0.257**	-0.179*	-0.098	-0.282***	1.000	
(7) Département rural	0.528***	-0.343***	0.588***	0.462***	0.430***	-0.252**	1.000

VIF = 2,5