



UNION NATIONALE DE L'AIDE, DES SOINS ET DES SERVICES AUX DOMICILES DU NORD



Impact social de la covid-19 : Décès et hospitalisations dans les services d'aide à domicile

L'impact social du coronavirus sur les
structures de la fédération UNA Nord

Arnaud BLANC

Encadré par Mme Lemaire, M. Loison et Mme Roussillon

2019/2020

Remerciements

Je tiens à remercier toutes les personnes qui m'ont aidé dans la réalisation de cette étude et au cours de ce stage.

Je remercie madame Roussillon, ma tutrice universitaire de l'Université Grenoble Alpes, de la faculté d'économie pour son aide au cours de la réalisation de cette étude et dans la rédaction du rapport de stage.

Je souhaite également remercier :

-Monsieur Thooft, président de la fédération UNA Nord, pour son accueil et son soutien tout au long de la réalisation de cette étude.

- madame Lemaire, directrice de la fédération UNA Nord et madame Bertin, assistante de direction, pour leurs aides dans la compréhension de ce secteur. En effet, elles m'ont donné accès à certaines données de contexte qui m'ont permis de mieux comprendre comment fonctionnait la fédération. Elles m'ont également apporté leur aide lors de la réalisation de l'enquête auprès des structures et lors de la rédaction de ce rapport.

- monsieur Loison, directeur de la Maison de l'Aide à Domicile de Lille, pour sa participation à cette étude et sa demande. Il m'a permis de présenter les résultats de cette étude à plusieurs occasions et a participé à l'organisation de l'enquête auprès des structures. Pour finir, je tiens à le remercier pour son aide dans la rédaction de ce rapport de stage.

-madame Lecyk, responsable projets à la Maison de l'Aide à Domicile, qui a pris le temps de répondre à toutes les questions que je me posais concernant le secteur de l'aide à domicile et qui m'a fait passer les documents dont j'avais besoin. Elle m'a également apporté son aide dans la rédaction de ce rapport.

-monsieur Six, Responsable du système d'information de la Maison de l'Aide à Domicile, qui m'a donné accès aux données de la Maison de l'Aide à Domicile de Lille. C'est aussi lui qui m'a permis de valider la méthodologie employée pour cette étude. Il m'a aussi apporté son soutien dans la rédaction de ce mémoire, en vérifiant la cohérence entre les résultats et la méthodologie employée.

Avant de terminer, je souhaiterais remercier les directeurs des structures qui ont pris le temps de répondre à l'enquête sur les décès et les hospitalisations.

Pour finir, j'aimerais remercier l'ensemble du personnel de la Maison de l'Aide à Domicile pour son accueil dans les conditions particulières qui ont marqué le début de ce stage. En effet, celui-ci a été marqué par le confinement national qui est intervenu le lendemain de mon arrivée.

Sommaire

I.	Introduction	7
II.	Revue de la littérature	9
1.	Les évolutions démographiques : une hausse du nombre de personnes âgées..	9
2.	Les caractéristiques des personnes bénéficiant de l'aide à domicile.....	10
3.	L'impact social du coronavirus sur la population : Une hausse des décès significative	10
4.	La méthodologie de l'évaluation d'impact	11
5.	Les facteurs explicatifs du recours à l'aide à domicile et des décès	12
III.	Le contexte de l'étude.....	16
1.	Une épidémie avec un impact social important.....	17
a)	L'évolution du nombre de décès dans le département du Nord	17
b)	Le coronavirus dans la région des Hauts-de-France	18
2.	Le secteur de l'aide à domicile.....	20
a)	Le marché de l'aide à domicile	20
b)	Les difficultés du secteur de l'aide à domicile en faveur des personnes âgées dépendantes.....	22
c)	Les difficultés du secteur de l'aide à domicile pendant le confinement.....	23
3.	Une fédération de l'aide à domicile : l'Union Nationale de l'Aide, des Soins et des Services aux Domiciles	24
a)	La fédération de l'UNA	24
b)	La fédération départementale l'UNA Nord	25
c)	Une association de la fédération : la Maison de l'Aide à Domicile de Lille ..	25
4.	Le modèle économique	26
a)	Les caractéristiques sociodémographiques.....	26
b)	L'état de santé des bénéficiaires de l'aide à domicile	27
IV.	Les données.....	28
1.	Les données mobilisées pour cette étude	28
a)	L'enquête effectuée auprès des structures.....	28
b)	Une deuxième source de données : l'INSEE.....	29
2.	Les variables utilisées pour les modèles	30
a)	Les variables utilisées pour les décès	30
b)	Les variables utilisées pour les hospitalisations	31
3.	L'analyse descriptive	32
a)	L'analyse descriptive sur l'échantillon des décès	32
b)	L'analyse descriptive sur l'échantillon des hospitalisations	35
V.	La méthodologie d'estimation	38
1.	La méthodologie d'estimation	39

2. La construction des modèles économétriques.....	40
a) Le modèle économétrique sur les décès	40
b) Le modèle économétrique sur les hospitalisations	42
VI. Les résultats	44
1. L'analyse des décès	44
a) Analyse de la robustesse des estimations par les Moindres Carrés Ordinaires (MCO)	45
b) L'interprétation des résultats obtenus par les Moindres Carrés généralisés (MCG).....	46
2. L'analyse des hospitalisations	47
a) Analyse de la robustesse des estimations par les MCO.....	49
b) L'interprétation des résultats par les MCG et par la correction de White.....	49
3. L'estimation des économies générées pendant la période de confinement....	50
VII. Conclusion.....	52
VIII. Bibliographie	53
IX. Annexes	57

Table des tableaux

Tableau 1 : Répartition des structures d'aide à domicile suivant le nombre de structures par type d'organisme, selon le ministère de l'économie des finances et de la relance	7
Tableau 2 : Variables explicatives potentielles pour expliquer le recours aux services d'aide à domicile et les décès	16
Tableau 3 : Répartition des heures rémunérées pour l'assistance aux personnes âgées et aux personnes en situation de handicap suivant le type d'organisme et le mode de prestation. Ainsi, les associations représentent 52.8% des heures rémunérées pour l'assistance aux personnes âgées en mode prestataire. Ces données sont calculées à partir de la répartition des heures d'intervention prestataire et mandataire par type d'activité en 2017, DARES.....	21
Tableau 4 : Tableau récapitulatif des différentes sources de données. <i>La colonne auteur correspond aux variables qui ont été construites par l'auteur de l'étude à partir des données provenant des 2 autres sources.</i>	30
Tableau 5 : Statistiques descriptives de l'échantillon de décès	33
Tableau 6 : Statistiques descriptives sur l'échantillon des individus hospitalisés	36
Tableau 7 : Résultats d'estimation par les moindres carrés ordinaires et par les moindres carrés généralisés.....	45
Tableau 8 : Résultats d'estimation sur les hospitalisations par les moindres carrés ordinaires et par les moindres carrés généralisés entre janvier 2019 et mai 2020 ...	48
Tableau 9 : Dictionnaire des variables de la base décès.....	59
Tableau 10 : Dictionnaire des variables de la base hospitalisations.....	60
Tableau 11 : Comparaison de l'échantillon des décès avant et pendant le confinement.....	61
Tableau 12 : Matrice des corrélations sur les décès.....	61
Tableau 13 : Comparaison de l'échantillon d'hospitalisations avant et pendant le confinement.....	62
Tableau 14 : Matrice des corrélations entre les variables des hospitalisations.....	63
Tableau 15 : Tableau d'analyse des variances sur les décès.....	63
Tableau 16 : p-value des tests de normalité	64
Tableau 17 : p-value du test de Breusch-Pagan de détection de l'hétéroscédasticité	65
Tableau 18 : Statistique de test du test de Durbin-Watson pour détecter l'autocorrélation	66
Tableau 19 : Test de multi-colinéarité entre les variables.....	67
Tableau 20 : Résultats des estimations des décès de la MAD depuis 2018.....	69
Tableau 21 : Comparaison des décès sur trois populations différentes	69
Tableau 22 : Comparaison de la population des personnes décédées dans le Nord et dans la MAD sur la population des plus de soixante ans	70
Tableau 23 : Test de multi-colinéarité entre les variables explicatives des hospitalisations	70
Tableau 24 : Tableau d'analyse des variances sur les hospitalisations.....	71
Tableau 25 : p-value des tests de normalité	72
Tableau 26 : p-value du test de Breusch-Pagan de détection de l'hétéroscédasticité	72
Tableau 27: Statistique du test de Durbin-Watson de détection de l'autocorrélation .	73
Tableau 28 : Estimation du nombre d'hospitalisations en niveau-niveau.....	74
Tableau 29 : répartition des individus selon le GIR dans les deux bases de données	74

Table des figures

Figure 1 : Nombre de décès par mois et années	17
Figure 2 : Evolution du nombre de décès rapporter à la densité de population.	18
Figure 3 : Poids des fédérations dans la prise en charge de la dépendance dans le département du Nord.....	22
Figure 4 : Evolution et moyennes mensuelles des décès.....	35
Figure 5: évolution des moyennes mensuelles d'hospitalisations.....	38
Figure 6 : QQplot des résidus.....	64
Figure 7 : Détection graphique de l'hétéroscédasticité	65
Figure 8 : Graphique après correction de l'hétéroscédasticité par la méthode des MCG	66
Figure 9 : Autocorrélogramme des résidus du modèle MCO sur les décès.....	67
Figure 10 : QQplot des résidus.....	71
Figure 11 : Détection graphique de l'hétéroscédasticité	72
Figure 12 : Graphique après correction de l'hétéroscédasticité par la méthode des MCG	73

Table des sigles

UNA : Union Nationale de l'Aide, des Soins et des Services aux Domiciles
INSEE : Institut National de la Statistique et des Etudes Economiques
PCH : Prestation de Compensation du Handicap
APA : Allocation Personnalisée d'Autonomie
GIR : Groupe Iso-Ressources
MAD : Maison de l'Aide à Domicile de Lille
EHPAD : Etablissement d'Hébergement pour Personnes Agées Dépendantes
ADMR : Aide à Domicile en Milieu Rural
DARES : Direction de l'Animation de la Recherche, des Etudes et des Statistiques
CNSA : Caisse Nationale de Solidarité pour l'Autonomie
CCAS : Centre Communal d'Action Sociale
MCO : Moindres Carrés Ordinaires
MCG : Moindres Carrés Généralisés
ATIH : Agence Technique de l'Information sur l'Hospitalisation
SAAD : Services d'Aide et d'Accompagnement à Domicile

I. Introduction

Le monde connaît une période de crise sanitaire importante due à l'apparition du Coronavirus. Il s'agit d'un virus proche de la grippe. Cette pandémie s'est déclarée à la fin de l'année 2019, en Chine. Ensuite, cette épidémie s'est rapidement propagée au reste du monde. En France, les premiers cas suspectés de coronavirus ont été détectés à Colmar, en novembre 2019. A partir du 14 février 2020, le coronavirus a provoqué un premier décès, en France. Cette crise a entraîné un confinement historique en France, dès le 17 mars. Celui-ci s'est étalé du 17 mars au 11 mai (Audureau, W. & Vaudano, M. 2020). Cette crise a eu de nombreuses conséquences économiques. En effet, pendant la période de confinement beaucoup d'entreprises ont fermé leurs portes. Néanmoins, d'autres activités ont continué. C'est le cas par exemple des activités d'aide à domicile. Ce secteur qui se situe habituellement dans l'ombre s'est retrouvé en pleine lumière. Les services d'aide à domicile ont continué à intervenir auprès des personnes âgées dépendantes, durant la période de confinement (Bretton, L. 2020).

Dans le même temps, la France connaîtra dans les années futures une importante augmentation du nombre de personnes âgées dépendantes. Ainsi, en 2070, l'INSEE estime qu'il y aura 8 millions de personnes de 75 ans ou plus. Ainsi, le nombre de personnes de 75 ans ou plus devrait être deux fois plus important qu'en 2013(INSEE, 2016).

Cette étude vise à évaluer l'impact social qu'ont eu les services d'aide à domicile pendant la crise. Celle-ci a permis de mettre en lumière un secteur qui demeurait dans l'ombre jusqu'à présent. En effet, le secteur de l'aide à domicile est un secteur complexe. Celui-ci est représenté par une multitude d'acteurs de tous types, aussi bien publics que privés. En effet, parmi ces acteurs on trouve des associations, des organismes publics comme les centres communaux d'action sociale ou encore des entreprises. Ainsi, en 2019, ce secteur était composé de 42147 organismes. Ceux-ci peuvent être répartis de la manière suivante :

Type d'organisme	Entreprises	Associations	Etablissements publics
répartition	81%	15%	4% ¹

Tableau 1 : Répartition des structures d'aide à domicile suivant le nombre de structures par type d'organisme, selon le ministère de l'économie des finances et de la relance

Le but de ces services est de faciliter la vie quotidienne des familles mais aussi d'accompagner les personnes fragiles, âgées ou en situation de handicap. Pour parvenir à cet objectif les services d'aide à domicile proposent certaines activités comme l'assistance aux personnes dépendantes, l'entretien de la maison et les travaux ménagers, la préparation ou la livraison des repas, la livraison des courses, le jardinage, le bricolage, les soins d'esthétique à domicile pour les personnes dépendantes, etc. L'assistance aux personnes dépendantes comprend notamment l'aide au lever et au coucher, à la toilette, à l'habillage et aux déplacements dans le logement.

¹ Ministère de l'économie et des finances, les chiffres clés du secteur des services à la personne : <https://www.servicesalapersonne.gouv.fr/donnees-et-etudes/chiffres-cles>

Ce secteur est également créateur de nombreux emplois. Il comptait 1.22 million de professionnels en 2016. En 2016, les entreprises de service d'aide à domicile ont connu une croissance de 8.5%. Pour finir, le secteur de l'aide à domicile a effectué plus de 861 millions d'heures rémunérées en 2016. Ce secteur est historiquement dominé par le milieu associatif qui reste encore majoritaire aujourd'hui. Dans les services d'aide à domicile, il existe aussi 2 modes de prestation : le mode prestataire où un organisme fait l'intermédiaire entre le client et le salarié et le mode mandataire où le client devient l'employeur du salarié. Le nombre d'heures en mandataire n'a cessé de baisser entre 2015 et 2018. Dans le cadre du mode prestataire, le nombre d'heures d'aide rémunérées est en hausse entre 2015 et 2018. Chaque année, il est possible d'observer une baisse des heures rémunérées pendant la période estivale. Le nombre d'organismes publics et d'entreprises privés est en hausse sur la même période (Ministère de l'économie et des finances, 2019).

Cette étude se concentrera sur les services d'aide à domicile qui interviennent dans le département du Nord. La mesure de cet impact se fera sur la base de 2 indicateurs principaux : le nombre de décès et le nombre d'entrées à l'hôpital. En effet, ceux-ci correspondent aux principaux chiffres qui sont sortis sur l'épidémie du Coronavirus. Il s'agit des principaux éléments qui mettront en avant l'impact des services d'aide à domicile.

Les services d'aide à domicile du réseau UNA Nord s'occupent majoritairement de personnes âgées dépendantes et de personnes en situation de handicap. En effet, le nombre d'heures consacrées aux personnes âgées dans les structures de la fédération UNA Nord étaient de 1 455 836 contre seulement 480 894 heures pour des personnes bénéficiaires de la PCH (Prestation de Compensation du Handicap), en 2018. Ainsi, la fédération UNA Nord s'occupait essentiellement de personnes âgées dépendantes qui représentaient 75.17%² de son activité, en 2018. Pour évaluer le nombre de décès qui ont pu être évités grâce aux services d'aide à domicile, il est possible d'utiliser la date de décès de chaque individu.

L'étude cherche à estimer le nombre de décès et d'hospitalisations qui ont été évités pour les personnes prises en charge par les services d'aide à domicile pendant le confinement. L'objectif sera également de mettre en avant les bénéfices ou les pertes associées à l'impact de l'aide à domicile sur les hospitalisations.

La base de données consacrée à cette étude est un panel d'individus qui sont morts entre le 1er janvier 2019 et le 31 mai 2020. Elle comporte le nombre de décès par mois qui ont eu lieu dans les différentes structures du réseau UNA Nord. Ce choix est fait puisque la base de données n'est composée que des données individuelles sur les personnes décédées et hospitalisées dans les structures adhérentes à la fédération UNA Nord.

Au niveau de chaque individu, les données comprennent l'âge, le genre, le GIR de l'individu. Des indicateurs sur la commune de décès sont aussi présents. Pour finir, ces bases de données sont composées de variables dichotomiques qui

² Bilan d'activité UNA Nord 2019 sur les chiffres de 2018

correspondent à la période de confinement. L'échantillon final qui porte sur les décès est composé de 620 individus.

Pour les hospitalisations, la base de données comprend le nombre d'hospitalisations, la durée des hospitalisations, l'âge de la personne hospitalisée, le genre de la personne, le Gir de la personne, la commune de la personne. Et une dummy correspondant à la période de confinement. L'échantillon portant sur les hospitalisations comporte 2808 individus qui sont entrées à l'hôpital.

Pour atteindre les objectifs de l'étude, la méthodologie qui sera mobilisée repose sur des données en coupe transversale. L'échantillon est composé de plusieurs coupes transversales à travers le temps qui sont agrégées.

Ainsi, les données disponibles reposent sur deux dimensions différentes : une dimension temporelle et une dimension individuelle. La dimension temporelle correspond à la date de décès et d'hospitalisations. La dimension individuelle représente les caractéristiques de chaque individu pris séparément. Ainsi, chaque échantillon est aléatoire. Les estimations seront donc effectuées par la méthode des moindres carrés ordinaires sur l'ensemble de l'échantillon.

Cette étude vise dans un premier temps à estimer l'impact social des services d'aide à domicile du réseau UNA Nord et de la Maison de l'aide à domicile (MAD). Puis dans un second temps, elle cherchera à monétariser leur impact social.

Pour cela cette étude débutera par une revue de la littérature. Puis elle enchaînera avec un retour sur le contexte. Celui-ci se fera à travers l'impact social du Coronavirus dans le Nord. Puis, le contexte du secteur de l'aide à domicile et le contexte de la fédération. Ensuite, elle continuera avec une présentation des données qui seront mobilisées dans cette étude. La troisième partie se consacrera à la méthode et à la construction du modèle. Pour finir, la dernière partie présentera les résultats de l'étude.

II. Revue de la littérature

1. Les évolutions démographiques : une hausse du nombre de personnes âgées

Dans les années à venir, la Population française va connaître de grands bouleversements. En effet, on assiste à un **vieillissement de la population**. Ainsi, en 2070, l'INSEE estime qu'il y aura 8 millions de personnes de 75 ans ou plus, en France. Cette hausse de la population concernerait en majorité les individus de 65 ans ou plus, soit environ 10.4 millions. Le nombre de personnes de 75 ans et plus devrait augmenter de 7.8 millions entre 2013 et 2070. En 2070, 13.7 millions d'individus seraient âgés de 75 ans ou plus. La proportion de 65 ans ou plus augmenterait fortement jusqu'en 2040. Ainsi, entre 2013 et 2040, la population résidant en France devrait augmenter de 8 points en 27 ans. Ensuite, entre 2040 et 2070, elle augmenterait seulement de 3 points. Cela s'explique par l'évolution des

personnes issues du baby-boom. En effet, cette population devrait atteindre l'âge de 65 ans ou plus d'ici 2039. Donc, la hausse des personnes âgées devrait ralentir après 2040. Le rapport entre le nombre de personnes de 65 ans ou plus et celui des 20-64 ans atteindrait 51%, en 2040. Il serait seulement de 57% en 2070, ce qui représente une hausse de 6 points entre 2040 et 2070. (INSEE, 2016)

Selon l'INSEE, on peut également observer que 4 millions de seniors seraient en perte d'autonomie en 2050. En 2015, en France, 2.5 millions de seniors étaient en perte d'autonomie. Parmi eux, 700 000 personnes peuvent être considérées en perte d'autonomie sévère. Parmi les seniors de 75 ans ou plus, 8.8% vivent en institution. Selon ce rapport, le nombre de places en hébergement permanent en établissements pour personnes âgées devrait augmenter de 20% d'ici à 2030 et de 50% à l'horizon 2050. La perte d'autonomie concerne en majorité les âges élevés. Ainsi, 30.2% des individus de 75 ans ou plus sont en perte d'autonomie, contre 6.6% des individus âgés de 60 à 74 ans. Dans le Nord, en 2015, entre 17 et 19% des personnes de 60 ans ou plus sont en perte d'autonomie. En 2015, seulement 8.8% des seniors de 75 ans ou plus vivaient en institution. Ce rapport fait aussi des prévisions pour 2050. Ainsi, la France contiendrait 4 millions de seniors en perte d'autonomie. Cela représenterait 16.4% des personnes âgées de 60 ans ou plus. Les personnes en perte d'autonomie sévère représenteraient alors 4.3% de la population des seniors. (INSEE, 2019)

2. Les caractéristiques des personnes bénéficiant de l'aide à domicile

La plupart des personnes âgées vivent à domicile. En effet, en 2016, 96% des hommes de 65 ans ou plus vivaient à domicile et 93% des femmes dans la même classe d'âge vivaient à domicile. Les femmes vivent en moyenne, plus longtemps que les hommes donc elles vivent aussi plus souvent seules. A l'inverse, les hommes vivent plus souvent en couple. Pour finir, l'hébergement en institution est plus fréquent chez les femmes que chez les hommes. Ce phénomène s'explique par l'espérance de vie qui est plus longue chez les femmes. La part de personnes admises en institution augmente avec l'âge à partir de 80 ans. L'âge d'entrée en institution est en moyenne de 85 ans et 2 mois en 2015. Le vieillissement augmente le risque de perte d'autonomie des personnes. La part des bénéficiaires de l'aide personnalisée d'autonomie augmente à partir de 75 ans. Ainsi, elle est de 6% pour les individus de 75 à 79 ans, 13% pour les 80 à 84 ans, 25% pour les 85 à 89 ans et 50% pour les 90 à 94 ans. Les plus de 95 ans sont 75% à bénéficier de l'APA (Allocation Personnalisée d'Autonomie). (INSEE, 2019)

3. L'impact social du coronavirus sur la population : Une hausse des décès significative

La crise actuelle liée à l'épidémie du coronavirus a impacté fortement la population. Au niveau national, le nombre de décès entre le 1^{er} mars et le 4 mai 2020 (129 678) est supérieur au nombre de décès recensé en 2018 (110 843) et 2019 (102 787). En effet, le nombre de décès est supérieur de 26% par rapport à 2019. Au niveau régional, on trouve que la région **Hauts-de-France est la troisième région en termes de décès**. En effet, elle en représente 27%.

Dans le Nord, le nombre de décès cumulés entre le 1^{er} mars et le 4 mai 2020 est supérieur de 22.7% par rapport à 2019. On remarque une forte baisse des décès entre le 1^{er} et le 15 avril et la période du 16 avril et le 4 mai. Cette baisse représente 26.7%. A l'inverse, durant la période du 16 au 31 mars et la période du 1^{er} au 15 avril 2020 le nombre de décès augmentait de 15.1%. En France, la hausse de la mortalité entre le 1^{er} mars et le 4 mai 2020 augmentait de 22% chez les femmes contre 23% chez les hommes, par rapport à 2019.

Dans la région Hauts-de-France, on trouve les résultats inverses. Les plus touchés par la mortalité sont des personnes qui ont entre **65 ans et 85 ans et plus**. La plupart des décès ont eu lieu à l'hôpital soit 50%. Ensuite, les décès ont lieu à domicile soit 24%. Les décès en EHPAD ne représentent que 15% des décès. Pour finir, 11% des décès ont lieu dans un lieu indéterminé. Les décès qui ont eu lieu à domicile entre le 1^{er} mars et le 4 mai, ont augmenté de 26% entre 2019 et 2020. 70% des personnes décédées ont entre 75 et 85 ans et plus. (INSEE, 2020).

D'après les données de santé publique France, les décès lié au coronavirus ont augmenté entre le 18 mars et le 31 mars. Après cette date, le nombre de décès a tendance à baisser.

Dans les Hauts-de-France, les personnes qui sont les plus touchées par la mortalité expliquée par le coronavirus, sont des individus qui sont plutôt âgés. Ainsi, les plus touchés sont les 80-89 ans qui représentent 40.44% des décès. Il y a ensuite, les 70-79 ans qui représentent 22% des décès du coronavirus. Pour finir, les données hospitalières montrent que les personnes ayant plus de 90 ans, représentent 19.78% des décès. Sur la période du 18 mars au 30 avril, dans le département du Nord, 111.6 hommes sont décédés contre 94.48 femmes. (Santé publique France, 2020)

4. La méthodologie de l'évaluation d'impact

Notre étude se concentre sur l'évaluation de l'impact de l'aide à domicile. Pour cela, il existe un certain nombre de méthodes. Ces pratiques cherchent à déterminer un effet causal de l'aide à domicile sur ces bénéficiaires. On distingue en général quatre méthodes de base. Celles ci sont les méthodes de l'expérience contrôlée, des variables instrumentales, de la double différence, de la régression sur discontinuité ou de l'appariement. (Givord, 2014).

Certaines études médicales s'intéressent à l'effet de l'aide à domicile sur différents bénéficiaires. Ainsi, un article de Sands étudie le volume de service à domicile qui cible les activités liées à la vie quotidienne, comme les services d'aide ménagère et les repas à domicile, par rapport au risque de transition des bénéficiaires. Pour cela, les auteurs utilisent une expérience contrôlée qui cherche à évaluer l'association entre le volume des services d'aide ménagère, les repas à domicile et la probabilité d'être placé en maison de retraite. Cette étude aboutit à un risque de placement en maison de retraite qui est significativement plus faible pour chaque augmentation de 5 heures dans les services d'aide à domicile (Sands, L.P. & al. 2012). L'étude de Rapp vise à déterminer si le recours à l'aide à domicile diminue la probabilité d'utiliser des soins d'urgence. Ils utilisent la méthode des variables

instrumentales. Ainsi, les bénéficiaires de l'aide à domicile ont un taux de soins d'urgence nettement inférieur aux non-bénéficiaires. En conclusion, l'aide à domicile a un effet protecteur pour les soins d'urgence (Rapp, T. & al. 2015).

Une étude de Citizing cherche à évaluer l'impact socio-économique de l'aide à domicile. Celle-ci se réfère à des travaux économétriques qui permettent d'établir des liens de causalité permettant d'identifier les paramètres qui ont un pouvoir explicatif et leur poids. Cette étude s'appuie sur des évaluations en différentiel. Dans ce cas, on compare une situation avec le programme d'aide à domicile appelé traitement à une situation sans l'aide à domicile, appelée contrôle. L'étude cherche à monétariser chaque impact en utilisant, les coûts d'opportunités, les valeurs tutélaires ou des études sur les consentements à payer. Elle se concentre sur plusieurs effets de l'aide à domicile identifiés.

Notre étude se concentrera sur l'aide à domicile apportée aux personnes dépendantes. Pour les personnes âgées dépendantes, il existe un lien de causalité entre l'aide à domicile et l'état général de ces personnes. Ainsi, une heure d'aide à domicile par semaine pendant un an diminue la probabilité de recourir aux urgences. (Rapp, T. & al. 2015). De la même façon, l'aide à domicile diminue la probabilité d'être contraint d'aller en maison de retraite. Pour finir, l'aide à domicile augmente la probabilité d'être en meilleure santé mentale et physique.

5. Les facteurs explicatifs du recours à l'aide à domicile et des décès

Pour étudier les décès, de nombreux facteurs peuvent être pris en compte. Ainsi, une étude sur les lieux de décès en France montre que le sexe, l'âge et le statut marital influent sur les lieux de décès. Ensuite, les auteurs utilisent un modèle de régression logistique pour déterminer les facteurs expliquant le choix du lieu de décès. Les facteurs retenus sont le sexe, l'âge, le statut matrimonial et les causes de décès. Le lieu de décès dépend significativement et positivement des dates. Il dépend aussi positivement de l'âge de l'individu. Il en est de même pour le statut matrimonial. (Aouba, A. Aubry, R. & al. 2012)

Certains facteurs ont tendance à augmenter la probabilité de décès. En effet, les personnes âgées cumulent des situations d'isolement et de dépendance. 90% des personnes âgées vivant à domicile bénéficient d'une aide à domicile. (Leblanc-Briot, M-T. 2014) D'autres études montrent que les **personnes âgées sont parmi les plus à risque** de décès lorsqu'elles sont contaminées par le coronavirus. Une note de santé publique France, du 18 mars rapporte que 1.3% des décès touchait les moins de 45 ans. Les plus de 75 ans représentaient 79.4% des décès. Dans cet article, on suppose que les décès liés au coronavirus à domicile, sont moins fréquents. **Les hommes sont légèrement plus touchés** que les femmes. Ainsi, Santé publique France indique qu'il y a 1.4 décès masculin pour un décès féminin.

Une étude a eu lieu au Québec. Elle montre que 75% des décès proviennent des centres d'hébergement, dans lesquels le virus se propage plus rapidement. Pour éviter les ravages d'une prochaine pandémie, il est possible de développer la demande pour les services d'aide à domicile. Le maintien des aînés dans leur domicile permettrait d'éviter ou de retarder l'hébergement en institution. Ainsi, le

maintien à domicile pourrait favoriser le bien-être des personnes âgées et générer des externalités positives. En effet, en cas de pandémie, la contagion sera limitée par rapport au milieu hospitalier ou les Etablissements d'hébergement pour les personnes âgées dépendantes. (Coulibaly, S. & al., 2020)

On cherche donc quels sont les déterminants du recours à l'aide à domicile. Pour cela, on va s'appuyer sur la littérature dont un article de Farajallah et Retali qui cherche à proposer une modélisation économétrique des variables qui influencent l'adoption des services d'aide au maintien à domicile. Dans ce cas, on retient en majorité des variables sociodémographiques.

Dans un premier temps, il semble que le genre est un indicateur important du fait de recourir ou non à un service d'aide à domicile. En effet, la majorité des études démographiques ont montré que les femmes ont une espérance de vie en moyenne, plus élevée que celle des hommes. Ainsi, à un certain âge les femmes ne peuvent plus recourir au soutien de leurs conjoints car elles sont devenues veuves. Cela augmente la probabilité d'utiliser une aide à domicile pour remédier à des difficultés physiques ou cognitives. L'âge est aussi une variable très importante. En effet, la vieillesse augmente la probabilité d'avoir recours à des services d'aide à domicile. En effet, l'augmentation de l'âge augmente le risque de dépendance et de maladie. Ce risque de dépendance oblige la personne âgée à faire appel à une aide.

Le fait d'être actif ou retraité pose aussi des questions. En effet, l'arrivée à l'âge de la retraite entraîne une baisse du revenu. Cette diminution oblige les retraités à réallouer une partie de leur budget. Ils vont chercher à baisser leurs dépenses de logement. Néanmoins, la personne âgée va également chercher à profiter au maximum de son domicile. Donc, elle va chercher à prendre un service d'aide à domicile plutôt que d'intégrer une maison de retraite. On peut aussi intégrer la solitude. En effet, se retrouver seul modifie le bien-être associé au logement. La solitude représente un coût pour la personne âgée, pouvant être réduit grâce à l'intervention d'un service d'aide à domicile. Le statut d'occupation du logement joue un rôle important sur la probabilité de recourir au service d'aide à domicile. En effet, les propriétaires ont plus de facilité pour aménager leur domicile, par exemple installer une chambre avec un lit médicalisé. A l'inverse, un locataire doit demander l'autorisation à son propriétaire pour mettre en place les aménagements dont il a besoin. D'ailleurs, il semblerait que les propriétaires aient plus recours à l'aide à domicile. Le niveau de revenu reste la variable déterminante pour accéder au service. Celui-ci décroît avec l'âge. Ainsi, une baisse du revenu devrait avoir un effet négatif sur le recours aux prestations d'un service d'aide à domicile. On pourrait dire la même chose du prix des services d'aide à domicile et notamment sur le reste à charge.

L'état de santé possède aussi un effet important sur l'adoption de l'aide à domicile. En effet, l'âge augmente la probabilité de devenir dépendant et donc de recourir à l'aide à domicile. La zone d'habitation est importante. En effet, les personnes âgées qui vivent dans des zones urbaines ont plus de facilité pour faire leurs courses puisqu'elles sont plus proches des commerces. A l'inverse, les personnes qui vivent en zone rurale sont plus isolées. Le lieu de résidence devrait avoir un effet positif sur l'utilisation d'un service d'aide à domicile. Le réseau familial

peut avoir un effet négatif sur la probabilité de recours au service d'aide à domicile (Farajallah, M. & Retali, G-R. 2013). L'ensemble de ces variables permettent de déterminer si une personne âgée fera appel à un service d'aide à domicile.

Variables explicatives possibles	Pourquoi ?	L'effet attendu sur l'utilisation des services d'aide à domicile	Indicateurs possibles
revenu	Le revenu est un élément important qui correspond aux ressources de la personne âgée. Cela peut représenter sa capacité à prendre soin d'elle. Ainsi, la personne âgée avec un revenu important peut investir pour faire en sorte de rester à domicile. On pourrait donc faire face à un biais de sélection positif. Néanmoins, il faut se souvenir que l'on travaille sur une structure du milieu associatif.	L'effet du revenu devrait être positif pour recourir à un service d'aide à domicile	-la catégorie socioprofessionnelle -Etre retraité ou non -l'âge -le genre -le fait d'être marié
Le prix	Il s'agit du prix du service d'aide à domicile. Ce prix est en parti pris en charge par les départements, par l'intermédiaire de l'APA. Dans le département du Nord, il s'agit d'une prise en charge de 21 €, pour chaque heure d'aide à domicile pour les plus bas revenus. Pour les personnes qui ont des ressources plus importantes, la prise en charge par le département est moins élevée. Celle-ci se situe entre 17 et 21€. Le reste doit être payé par la personne âgée. Or,	L'effet devrait être négatif	

	cette personne voit son revenu baisser. Donc, ce reste à charge peut-être dissuasif pour la personne âgée.		
L'état de santé	L'état de santé est un élément essentiel du choix de recourir à un service d'aide à domicile. En effet, une personne âgée qui a de gros problèmes physiques ou psychiques et qui ne peut plus se déplacer aura recours à un service pour faire ses courses, pour faire sa toilette, pour faire son ménage et pour préparer ses repas, etc. Le niveau de dépendance est mesuré par le GIR. Il est possible d'avoir un biais de sélection négatif. En effet, si on part de l'éventualité que les personnes qui sont les plus dépendantes font appel au service d'aide à domicile alors elles ont un risque plus important d'aller à l'hôpital ou en maison de retraite qu'une personne moins dépendante qui n'aurait pas recours à l'aide à domicile. On risque donc, de sous estimer l'effet de l'aide à domicile.	Le niveau de dépendance devrait avoir un effet positif sur le recours à un service d'aide à domicile	-le GIR -la durée de l'aide à domicile -Le fait de souffrir d'une maladie chronique -âge
Les substituts	La personne âgée peut remplacer un service d'aide à domicile par un aidant familial si la prestation du service est trop coûteuse. Cela est vrai aussi	Le fait que le conjoint soit toujours en vie ou que le couple ait des enfants à proximité diminue le recours à un service d'aide à domicile	-Conjoints

	<p>tant que la personne âgée a un proche en capacité d'effectuer les tâches courantes. On remarque, néanmoins, que les enfants ont tendance à moins prendre en charge leurs parents (Fontaine, R. & Arnault, L. 2016)</p>		
La temporalité de l'étude	<p>La saison peut avoir un impact marqué sur le nombre de décès. En effet, l'été est souvent marqué par des phases caniculaires. Dans ce cas, le nombre de décès des personnes bénéficiaires de l'aide à domicile devrait augmenter. De la même façon, l'hiver est une période marquée par une abondance d'épidémies (grippe...). On devrait donc observer une recrudescence des décès pendant l'hiver</p>	<p>Ainsi, le fait d'être dans un mois d'hiver devrait entraîner une augmentation du nombre de décès parmi les bénéficiaires des services d'aide à domicile</p>	-les saisons de décès
Le Coronavirus	<p>Le coronavirus a entraîné une surmortalité, en France pendant la période de confinement (INSEE, 2020)</p>	<p>Le coronavirus devrait donc entraîner une augmentation du nombre de décès</p>	-Le confinement (mars à mai 2020)

Tableau 2 : Variables explicatives potentielles pour expliquer le recours aux services d'aide à domicile et les décès

III. Le contexte de l'étude

Cette étude se développe dans un contexte très particulier puisqu'elle se concentre sur la période du confinement. Durant cette phase, les services d'aide à domicile ont été mis en lumière du fait de leur rôle essentiel pour le maintien à domicile des personnes âgées et des personnes en situation de handicap.

De plus, cette étude se limite à une dimension géographique. En effet, cette étude se limite au département du Nord. Ainsi, Cette partie se décomposera en quatre sous parties. La première se concentrera sur le confinement et sur l'impact de ce confinement. La deuxième partie se concentrera sur le secteur de l'aide à domicile

et la suivante se focalisera sur l'une des fédérations d'aide à domicile du département du Nord, UNA Nord. Pour finir, la dernière partie vise à définir les hypothèses testées du modèle économique.

1. Une épidémie avec un impact social important

a) L'évolution du nombre de décès dans le département du Nord

Cette épidémie a eu un impact important, notamment au niveau social. Elle a entraîné un nombre de décès conséquent. On peut remarquer que le nombre de décès pendant la période du confinement a été significativement plus important que le nombre de décès qui a eu lieu à cette même période les années précédentes. Ainsi, on obtient le graphique suivant à l'échelle du département du Nord :

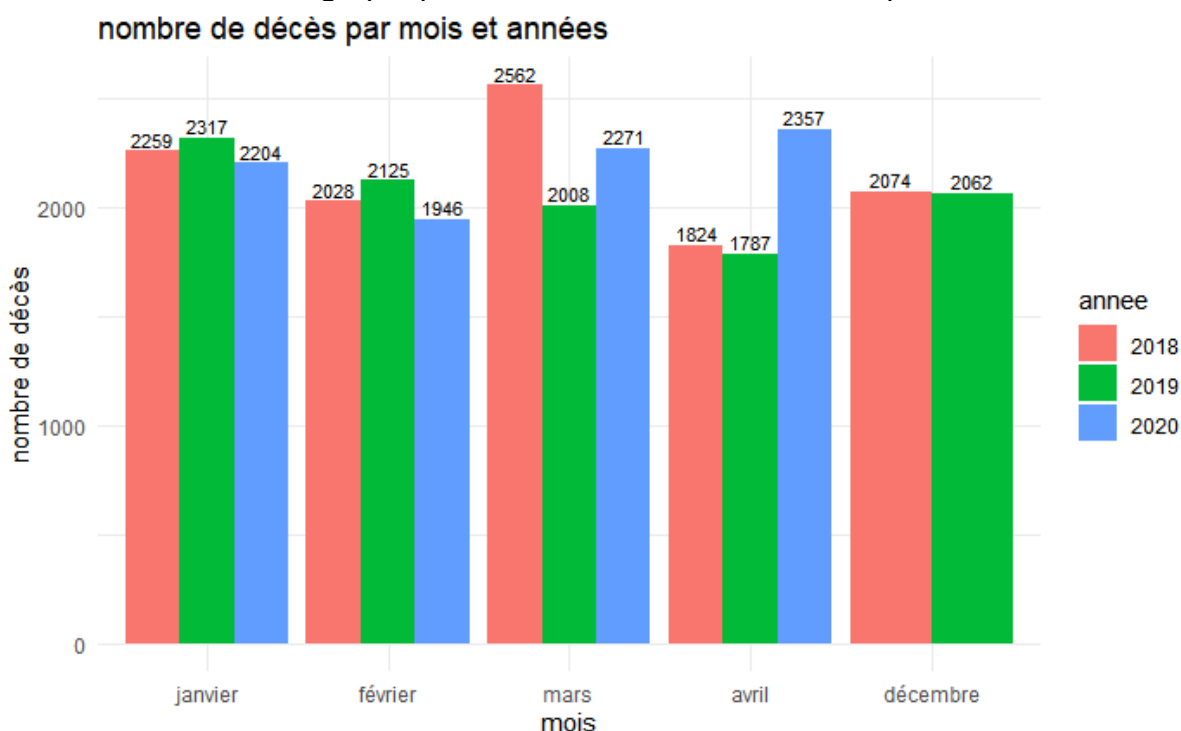


Figure 1 : Nombre de décès par mois et années, données : INSEE provenant des services de l'Etat-civil des communes

On remarque que le nombre de décès a fortement augmenté en avril. En effet, le nombre de décès du mois d'avril 2020 est supérieur au décès des années précédentes. De plus, on remarque que le nombre de décès pendant la période de confinement est significativement supérieur au nombre de décès à la même période les années précédentes. Il y a donc eu une surmortalité pendant le confinement.

En effet, en moyenne, **durant la période de confinement**, le nombre de décès était de 75.869 décès par jour contre 62.213, en 2019. Donc, les décès ont augmenté de 13.656. Cela représente **une augmentation de 21.95%**. De la même façon en 2018, en moyenne, pendant la période de confinement, le nombre de décès étaient de 71.902. Le nombre de décès a ainsi augmenté de 3.967 soit une hausse de 5.52%. La période de **confinement** est donc une **période de surmortalité** par rapport aux années précédentes. Cette surmortalité dépend fortement de la

commune du Nord dans laquelle on se situe. En effet, comme il est possible de le voir sur la carte suivante, certaines communes ont connu une forte hausse des décès par rapport à la densité de population vis-à-vis d'autres communes de ce département.

Evolution du nombre de décès ramener à la densité de population

1 part des individus décédés/km² (%)

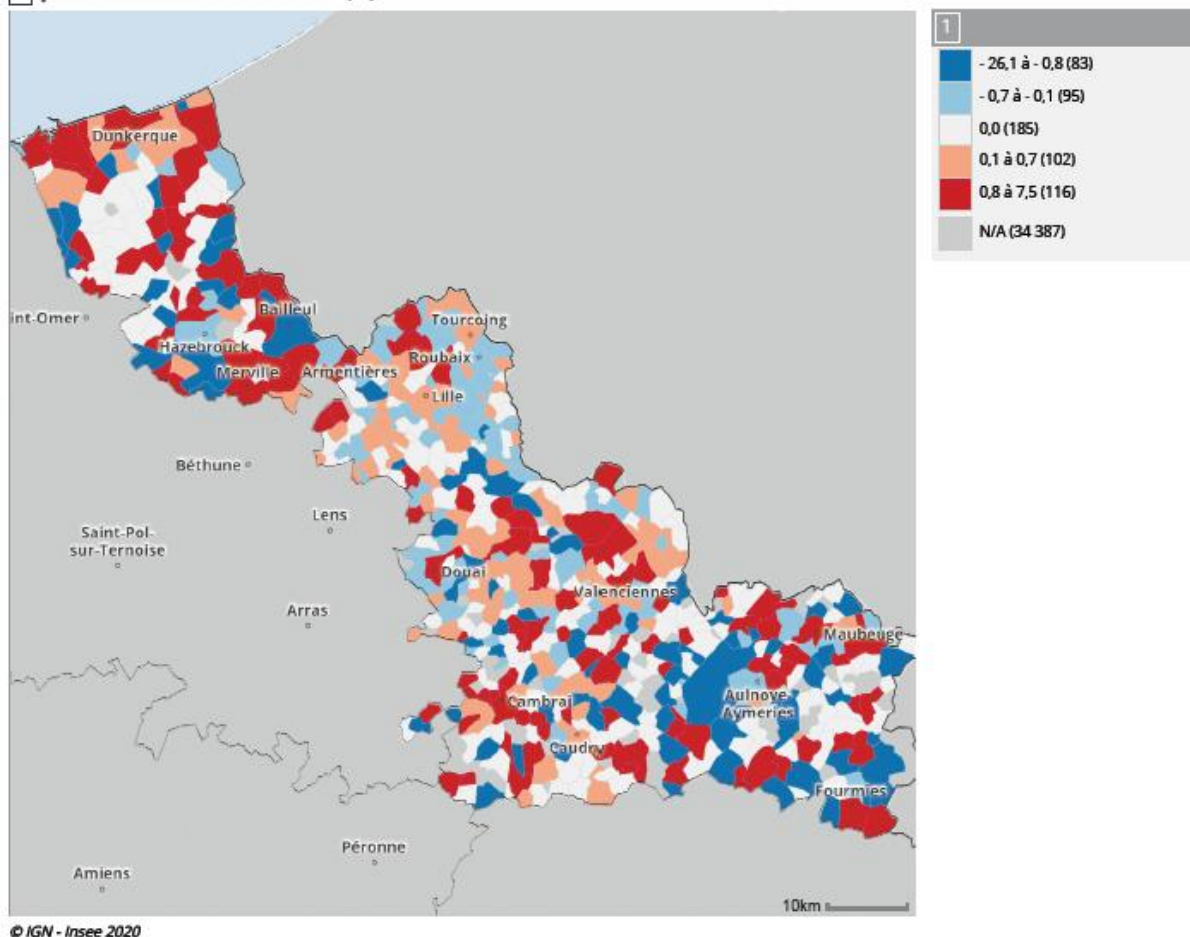


Figure 2 : Evolution du nombre de décès rapporté à la densité de population.

Cette carte donne l'évolution du nombre de décès par commune pour 100 habitants/km², données : INSEE

On peut remarquer grâce à cette carte que les communes qui ont connus une **hausse des décès** entre mars et mai 2020 par rapport à mars et mai 2019 sont des communes avec une **densité de population élevée**. En effet, cette carte représente l'évolution des décès rapportée à la densité de population. Ainsi, dans une ville comme Dunkerque, le nombre de décès a augmenté de 70 entre 2019 et 2020. Cette augmentation représente 3.5 décès en plus pour 100 habitants/km². Il existe quelques exceptions. C'est le cas par exemple de Roubaix où le nombre de décès pour 100 habitants/km² a baissé de 0.3.

b) Le coronavirus dans la région des Hauts-de-France

Le premier cas de Coronavirus détecté dans les Hauts-de-France date du 26 février, dans l'Oise (ARS des Hauts-de-France, 2020). Dans la région des Hauts-de-France, le pic du nombre de cas a été enregistré fin mars 2020. Ensuite, il a commencé à décroître avec la mise en place du confinement. On assiste bien à une

baisse du nombre de consultation pour infection respiratoire, à la suite du confinement.

Dans les EHPAD, on remarque que le nombre de coronavirus signalés est de 7779, le 28 mai 2020. Parmi ces cas, il y a eu 675 décès en institutions et 510 décès à l'hôpital. Selon Santé Publique France et l'INSEE, on remarque un excès de mortalité au niveau régional. Le pic de mortalité apparaît entre la fin mars et le début avril. Néanmoins, dans les semaines qui suivent le confinement il n'y a plus de surmortalité dans la région des Hauts-de-France, dont fait partie le département du Nord.

Dans la région des Hauts-de-France, 1712 cas ont été confirmés et 2070 sont des cas possibles. Le nombre de cas a légèrement augmenté après le confinement dans le département du Nord. La mortalité qui est spécifique au coronavirus, c'est-à-dire les décès des patients hospitalisés pour coronavirus et en institution. Entre le 1^{er} mars et le 27 mai 2020, Santé Publique France rapporte 2391 décès. Parmi ces décès, il y a eu 1716 décès chez des patients hospitalisés contre 675 décès en établissements. Ainsi, le nombre de décès en établissement représentait 28.23%. La plupart des personnes qui sont décédées à l'hôpital du coronavirus étaient âgées de plus de 70 ans. Ces personnes représentaient plus de 81% des décès des personnes hospitalisées.

Une autre source est utilisée pour surveiller la mortalité. On parle alors de mortalité non spécifique. Cette source s'appuie sur les données transmises par les services de l'état-civil des communes. Ces données prennent en compte 80% de la mortalité de la région des Hauts-de-France. Ces données ne fournissent aucune information sur les causes de décès. Parmi ces données, on observe un pic de mortalité la semaine du 6 avril 2020. Cela correspond à la semaine qui suit celle du pic expliqué par le coronavirus. Parmi les décès du coronavirus chez les personnes hospitalisées, on voit que les personnes les plus touchées sont âgées de 80-89 ans. La plupart des décès touchent des personnes qui ont plus de 60 ans (Degrenel, M. & al. 2020). Le coronavirus a donc eu un impact social important puisque le nombre de décès a beaucoup augmenté durant la période du confinement.

Du point de vue des hospitalisations, les évolutions sont également importantes. En effet, d'après le point épidémiologique régional du 28 mai 2020, Le nombre d'hospitalisations liées au coronavirus touche majoritairement des personnes âgées. Ainsi, 62% des individus de plus de 65 ans sont hospitalisés après un passage aux urgences pour suspicion de coronavirus. Cette tranche d'âge est également majoritaire dans les services de réanimations (75%). Le 24 mai 2020, il restait 1679 personnes atteintes du coronavirus dans les hôpitaux. Au total, depuis le 1^{er} mars 2020, 9279 patients ont été hospitalisés pour cause de coronavirus. Parmi ces patients 35% ont été pris en charge dans le département du Nord. Au plus haut de l'épidémie, durant la semaine du 23 au 29 mars, le coronavirus représentait environ 20% de l'activité hospitalière. Par la suite, la part de l'activité hospitalière dédiée au coronavirus n'a cessé de baisser pour atteindre environ 2% entre le 18 et le 24 mai 2020. Cela représente une baisse de 18 points entre le pic de l'épidémie et le 24 mai 2020. (Degrenel, M. & al. 2020).

Pendant la période de confinement, l'activité des médecins et certaines activités hospitalières hors de la prise en charge du coronavirus ont fortement baissé. Ainsi, l'assurance maladie comptabilisait une baisse de la fréquentation des cabinets de médecins de près de 40%. De la même façon, les consultations chez des spécialistes ont baissé de 50% pendant le confinement lié à la crise sanitaire. En effet, certains patients à risques ont cessé de consulter pendant la crise du coronavirus, comme des patients diabétiques, des malades chroniques, cancéreux ou insuffisants cardiaques.

De la même façon, on assiste à une baisse des fréquentations des hôpitaux pendant la crise sanitaire. En effet, l'Assistance publique-Hôpitaux de Paris remarque une baisse de certaines prises en charge. Ainsi, les cas d'appendicites ont baissé de 74% et les cholécystites ont baissé de 47%. (Godeluck, S. 2020). Selon le journal la voix du Nord, l'hôpital d'Arras a connu une forte baisse d'activité pendant la crise du coronavirus. En effet, entre le 13 mars et le 13 avril 2020, le nombre de passages aux urgences a baissé de 50%. Il y a eu 831 hospitalisations hors Covid. Cela représente une baisse de 25% par rapport à 2019. Dans le même temps, le nombre de consultation a baissé de 75%. Le nombre de journées en réanimation a atteint 896 entre le 13 mars et le 13 avril soit une multiplication par deux par rapport à la même période en 2019. (La voix du nord. 2020)

2. Le secteur de l'aide à domicile

a) Le marché de l'aide à domicile

Le secteur de l'aide à domicile est un secteur complexe qui regroupe une multitude d'acteurs. En effet, le secteur de l'aide à domicile est composé à la fois d'organismes publics, d'associations et d'entreprises privées depuis le plan Borloo de 2005. Celui-ci a favorisé l'ouverture du marché des services à domicile aux entreprises privées lucratives. Le but recherché était de lutter contre le chômage en créant des emplois (Duthil, G. 2007). En 2019, le secteur de l'aide à domicile comptait 42 147 organismes. Parmi ces organismes, on trouve une majorité d'entreprises privées. Ce secteur est donc devenu de plus en plus concurrentiel. Selon Gilles Duthil, en 2007, la fédération de **l'UNA représentait 14.5%** des parts de marché, l'ADMR représentait 11% des parts de marché et ADESSA représentait 3.5%. Ainsi, le secteur de l'aide à domicile était dominé par les associations. En 2018, **l'activité des associations** d'aide à domicile reste **majoritaire**. En effet, les associations représentaient **49.8% des heures rémunérées** contre 41.4% pour les entreprises privées. Les derniers acteurs de ce marché sont les organismes publics qui représentaient 8.8% des heures rémunérées, en 2018 (DARES, 2020).

Le secteur de l'aide à domicile peut aussi se décomposer en plusieurs branches d'activités. En effet, on peut retenir trois principaux éléments : les services de la vie quotidienne, les services aux publics fragiles et les services aux familles. Le premier élément comprend le ménage et l'entretien de la maison, les travaux d'entretien du jardin, des travaux de bricolage, la préparation de repas à domicile, la collecte et la livraison de linge repassé, la livraison des courses à domicile. Ensuite, il y a les services aux publics fragiles. Ceux-ci comprennent l'assistance aux personnes âgées ou à toutes autres personnes qui auraient besoin d'une aide

personnelle, l'assistance aux personnes en situation de handicap, la conduite du véhicule des personnes dépendantes, l'aide à la mobilité pour les personnes dépendantes, les soins d'esthétique à domicile et les soins et promenades d'animaux de compagnie. Pour finir, il existe une troisième branche de service qui est les services aux familles. En plus de ces trois branches d'activités, on trouve 2 grands types de modes de prestations de service : le mode prestataire et le mode mandataire. Avec le mode prestataire, le particulier aidé achète à un organisme une prestation qui lui est facturée. Avec le mode mandataire, le particulier devient l'employeur de l'intervenant à domicile. L'organisme n'a qu'une fonction de ressources humaines. En 2018, les services prestataires connaissent une hausse de leur activité de 1.1% par rapport à 2017 (DARES, 2020).

L'activité principale du secteur est l'assistance aux personnes âgées. En effet, cette activité représentait **43.8% des heures d'intervention**, en 2017. C'est aussi la principale activité des associations qui étaient dominantes dans l'ensemble des activités prestataires, selon les heures rémunérées. Les associations consacraient, en effet, 52.8% de leurs heures rémunérées à l'assistance aux personnes âgées contre seulement 31.3% pour les entreprises en mode prestataire. En mode mandataire, la répartition est légèrement différente. En effet, en 2017, la répartition du nombre d'heures rémunérées se fait majoritairement en faveur des personnes âgées dans les associations (47.1%) et dans les organismes publics (47.7%). A l'inverse, dans les entreprises c'est l'entretien de la maison qui domine. En effet, cette activité représente 34% de l'activité des entreprises mandataires contre 24.2% pour l'assistance aux personnes âgées. (DARES, 2020)

Type d'organisme	Entreprises		Associations		Organismes publics	
Répartition	81%		15%		4%	
Assistance aux personnes âgées	prestataire	mandataire	prestataire	mandataire	Prestataire	mandataire
	31.3%	24.2%	52.8%	47.1%	48.3%	47.7%
Assistance aux personnes en situation de handicap ³	Prestataire	Mandataire	Prestataire	Mandataire	Prestataire	Mandataire
	8.1%	3.4%	9.1%	9.9%	3.7%	5.6%

Tableau 3 : Répartition des heures rémunérées pour l'assistance aux personnes âgées et aux personnes en situation de handicap suivant le type d'organisme et le mode de prestation. Ainsi, les associations représentent 52.8% des heures rémunérées pour l'assistance aux personnes âgées en mode prestataire. Ces données sont calculées à partir de la répartition des heures d'intervention prestataire et mandataire par type d'activité en 2017, DARES.

Le graphique suivant confirme les résultats trouvés au niveau national sur le département du Nord :

³ Il s'agit de la part des heures rémunérées

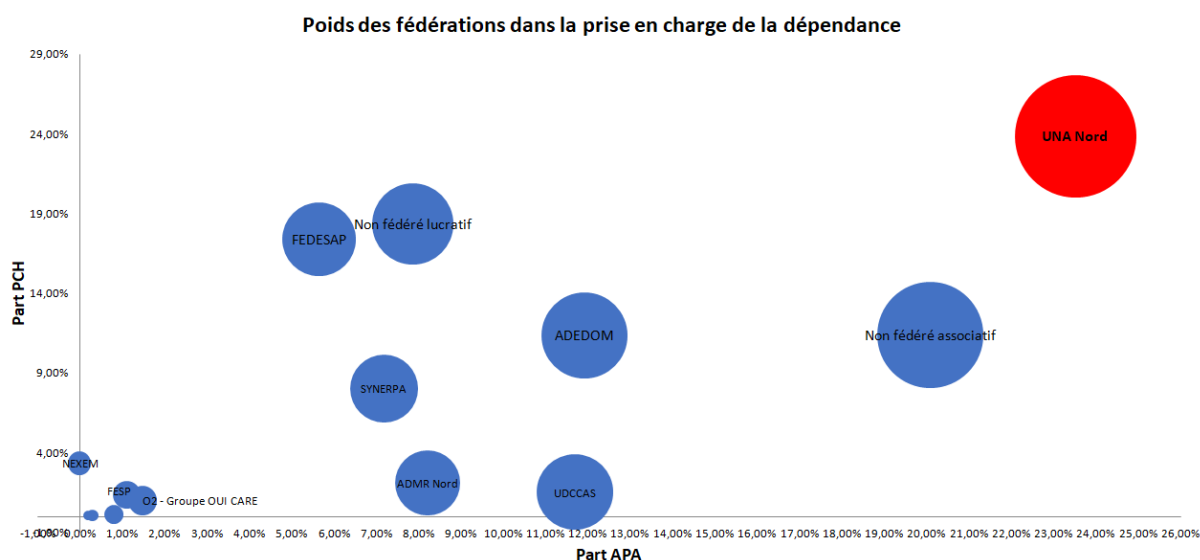


Figure 3 : Poids des fédérations dans la prise en charge de la dépendance dans le département du Nord. Ce graphique représente le nombre d'heures d'aide à domicile consacrées aux personnes âgées dépendantes (APA) et le nombre d'heures consacrées aux personnes en situation de handicap (PCH). La taille des bulles correspond au nombre d'heures totales pour chaque fédération

Ainsi, on peut remarquer que le nombre d'heures est majoritaire dans les fédérations associatives. Dans le Nord, **la fédération qui prend en charge le nombre d'heures totales le plus important** est l'union nationale de l'aide, des soins et des services aux domiciles du Nord (**UNA Nord**). Cette fédération d'associations prend aussi bien en charge des personnes en situation de handicap (23.84%) que des personnes âgées dépendantes (23.51%). On peut aussi remarquer que les fédérations d'associations et les organismes publics prennent surtout en charge des personnes âgées dépendantes. C'est le cas par exemple, d'ADEDOM (11.8%), de l'UDCCAS (11.5%) ou de l'ADMR du Nord (8.3%). Les fédérations d'entreprises prennent en charge des individus qui sont majoritairement en situations de handicap. En effet, la fédération FEDESAP et les non fédérés lucratifs représentent plus de 17% des parts PCH. A l'inverse, elles prennent en charge moins de personnes âgées dépendantes.

Pour finir, les données disponibles montrent que le secteur associatif consacre plus de 6 millions d'heures rémunérées à la dépendance. Les fédérations du secteur privé de l'aide à domicile ne représente qu'environ 2 millions d'heures. Ainsi, au niveau de la dépendance, **les fédérations d'associations et d'organismes publics représentent environ 72 % des heures rémunérées contre seulement 28% pour les fédérations d'entreprises dans le département du Nord, en 2020.**

b) Les difficultés du secteur de l'aide à domicile en faveur des personnes âgées dépendantes

Néanmoins, les services d'aide à domicile ont des difficultés économiques importantes. Celles-ci s'expliquent notamment par une **tarification** qui est en moyenne de **21.67 €**, en France pour l'APA. Ce tarif est inférieur au **coût de revient** qui avait été évalué en moyenne à **24€**, en 2016 (CNSA- DGCS, 2016). De plus l'offre est atomisée sur toute la France. Ainsi, on dénombre 7 000 services en France. Les services dépendent en majorité de la branche de l'aide à domicile qui représente 60% des services d'aide à domicile. Mais, ces services peuvent aussi

dépendre d'autres branches comme celles des services à la personne qui représentent 20% des services d'aide à domicile. Il s'agit en majorité d'entreprises privées lucratives. Pour finir, on trouve également des services d'aide à domicile qui dépendent de la fonction publique territoriale comme les centres communaux d'action sociale (CCAS) qui représentent environ 10% des services (Libault, D. 2019).

L'activité de l'assistance aux personnes âgées est aussi marquée par des disparités territoriales importantes. Ainsi, l'allocation personnalisée d'autonomie varie selon les départements. Ainsi, d'après une enquête de la CNSA auprès de 88 conseils départementaux les **pratiques tarifaires** vis-à-vis des services d'aide à domicile **sont très diverses**. Ainsi, en 2016, la médiane des tarifs dépendance des établissements, fixés par le conseil départemental, s'établit à 5.5€/jour pour les GIR 5 et 6, 13€/jour pour les GIR 3 et 4 et 20.4€/jour pour les GIR 1 et 2. Néanmoins, les fluctuations de ces tarifs varient de façon importante entre les départements. Les prestations de services et leurs prises en charges sont très complexes à comprendre. En effet, elles font intervenir une multitude d'acteurs. Ainsi, les dépenses d'aide à domicile sont financées en partie par les conseils départementaux et les caisses de retraite. En effet, les conseils départementaux prennent en charge la dépendance et l'aide à la vie quotidienne par l'intermédiaire de l'allocation personnalisée d'autonomie à domicile. De plus, comme on l'a vu précédemment, le bénéficiaire fait face à une multitude d'acteurs aux statuts différents (Libault, D. 2019).

L'aide à domicile fait aussi face à une troisième difficulté majeure. Cette difficulté est liée à **l'emploi**. En 2018, il y avait 270 000 équivalents temps plein dans les services d'aide et d'accompagnement à domicile (Dares, 2018). Les salariés qui travaillent dans des services d'aide à domicile sont plutôt âgés. En effet, l'âge moyen des salariés est proche de 43.6 ans. En plus de l'âge élevé des salariés, les employeurs ont aussi des **difficultés de recrutement**. Ainsi, 17 000 postes seraient à pourvoir immédiatement selon le rapport de la branche de l'aide à domicile de 2017. De plus, 19% des postes créés en 2018 seraient restés vacants selon l'UNA. De plus, les besoins risquent de fortement augmenter dans les années qui viennent. En effet, on assiste à un vieillissement de la population française avec une augmentation des personnes âgées dépendantes, selon l'INSEE. Ainsi, les besoins en emploi d'ici 2030 atteindraient entre 150 et 200 000 équivalents temps plein, selon France Stratégie et la DARES (Libault, D. 2019).

c) Les difficultés du secteur de l'aide à domicile pendant le confinement

Le secteur de l'aide à domicile comme tous les secteurs de l'économie a souffert du confinement. Néanmoins, les salariés du secteur de l'aide à domicile **ont continué à intervenir** sur le terrain. Cependant, ils étaient bien démunis au début du confinement. En effet, avant même le confinement, les services d'aide à domicile ont souffert du **manque de masques** (Steg, P. G. 2020). Les soignants à domicile ont eu beaucoup de mal à obtenir les équipements de protection car ils n'étaient pas considérés comme prioritaires. En effet, lors de la période de confinement des milliers de salariés de l'aide à domicile ont continué à intervenir au domicile des

personnes âgées ou en situation de handicap, directement au contact des personnes. En effet, ils effectuent des aides au lever, à la toilette, au repas et au coucher. Ces activités sont primordiales pour permettre aux personnes dépendantes de rester à domicile. Or, les aides à domicile n'ont eu accès que tardivement aux équipements de protection.

De plus, les salariés des structures d'aide à domicile ont eu du mal à avoir accès à la garde d'enfants. Certaines auxiliaires de vie se sont retrouvées en arrêt de travail ou ont été obligées de garder leurs enfants. D'autres salariés ont eu peur d'apporter le virus chez les patients ou bien de contaminer leur famille. Ainsi, l'ADMR travaillait avec seulement 60% de ses effectifs (Silbert, N. 2020). Au début de cette crise, les prestations de ménage ont dû être suspendues (Meslet, E. 2020). Au 17 mars 2020, certaines associations d'aide à domicile avaient mis en place le chômage partiel. De plus certains salariés ont fait valoir leur droit de retrait. (hospimédia, 2020) Un autre article d'hospimédia met aussi en relief le manque de reconnaissance envers ce secteur pendant la crise du coronavirus. En effet, les représentants de la branche de l'aide à domicile déplorent le fait que les salariés du secteur soient considérés comme des acteurs secondaires. Ainsi, ils se sont vus refuser la délivrance des masques en officines. (hospimédia, 2020)

Malgré ces problèmes importants de manque de moyens de protections, les salariés de l'aide à domicile ont continué à prendre en charge leurs bénéficiaires. De plus, les services d'aide à domicile ont été encore **plus sollicités**. En effet, ceux-ci ont dû prendre en charge des personnes âgées qui sortaient des hôpitaux et les maintenir à domicile. Ce maintien à domicile a permis aux hôpitaux de **libérer des lits** pour recevoir les patients atteints du coronavirus.

3. Une fédération de l'aide à domicile : l'Union Nationale de l'Aide, des Soins et des Services aux Domiciles

a) La fédération de l'UNA

Il s'agit d'une association à but non lucratif qui fut fondée en 1970. Cette association fédère les associations du secteur de l'aide à domicile. Elle est réparties en 21 unions régionales et 55 unions départementales ou interdépartementales. Ainsi, cette fédération qui emploie **75 718 salariés, regroupe plus de 732 structures** à travers la France, en 2018. Ces structures sont en majorité des associations et des services publics territoriaux. Celles-ci gèrent plusieurs types de service auprès de différents publics. Ainsi, certaines structures sont spécialisées dans l'aide à domicile auprès des personnes âgées, dans l'aide aux familles et d'autres missions de soins à domicile. Elles sont réparties de la façon suivante :

- Il y a 619 associations soit 84.56%
- On retrouve 106 CCAS soit 14.48%
- On a également 7 mutuelles soit 0.96%

Les activités de cette fédération sont tournées vers les publics fragiles. Ainsi, en 2018, le réseau a effectué 67 349 562 heures auprès des publics fragiles. La part la plus importante des heures est destinée aux personnes âgées. En effet, les **personnes âgées représentent 64% des heures effectuées** par le réseau. La

majorité des structures de l'UNA se consacre aux services aux personnes âgées (89.2%). Le deuxième public le plus représenté est composé des personnes en situation de handicap. En effet, 88.7% des structures prennent en charge des personnes en **situation de handicap**. Cette mission ne représente que **10% des heures enregistrées en 2018**.⁴

b) La fédération départementale l'UNA Nord

La fédération départementale de l'UNA Nord comprenait **26 structures** adhérentes, dont **23 associations** et 3 CCAS, en 2018. Parmi ces structures, on retrouve 17 structures prestataires et 10 structures qui fonctionnent en mode mandataire. Certaines structures proposent les différents modes de prestations. Les structures adhérentes comprenaient 10 services de soins infirmiers et 2 services familles. Ces services ne seront pas étudiés dans cette recherche. Ainsi, les adhérents employaient **3 820 salariés** au 31 décembre 2018. Ces salariés sont intervenus auprès de **20 550 personnes fragiles**. Depuis 2007, on assiste à une baisse de l'activité du réseau. Comme expliqué précédemment, le mode prestataire s'accroît depuis 2016. Le mode mandataire connaît une baisse continue depuis 2007.

Ce réseau est porté par une forte activité prestataire en faveur des personnes âgées dépendantes. En effet, en 2018, les personnes qui avaient un niveau de GIR compris entre 1 et 4 représentaient 52.49% de l'activité des adhérents et les personnes fragiles en situation de handicap représentaient 17.34% des personnes prises en charge. Ainsi, la majorité des heures étaient réalisées au bénéfice des personnes âgées dépendantes soit 1 455 836 heures contre seulement 480 894 heures destinées aux personnes en situation de handicap. En 2018, le nombre d'heures dédiées à un bénéficiaire de l'APA représentait 49.61% des heures effectuées par les adhérents, pour tous les modes de prestation. A l'inverse, les heures dédiées aux services aux personnes en situation de handicap qui bénéficient de la PCH ne représentaient que 16.39% de l'activité des structures adhérentes de l'UNA.⁵

c) Une association de la fédération : la Maison de l'Aide à Domicile de Lille

La Maison de l'Aide à Domicile de Lille est l'une des associations d'aide à domicile qui a adhéré à la fédération UNA Nord. Cette structure se compose de 2 associations adhérentes d'aide à domicile. Ces deux associations sont l'ASSAD et l'AMAGAD. La maison de l'aide à domicile agit dans les communes de la métropole de Lille. Elle intervient auprès de tous les publics, mais en majorité auprès des personnes âgées et des personnes en situation de handicap. Dans ce cas, le but de l'aide à domicile est de permettre à la personne de demeurer à son domicile le plus longtemps possible et de prolonger au maximum son autonomie. Les personnes en situation de handicap forment la seconde activité principale de la Maison de l'Aide à Domicile. Cette structure comptabilisait à elle seule plus de **5074 clients** au 31

⁴ Bilan d'activité de la fédération UNA national(2018). [<https://capatools.fr/una-rapport/#2>]

⁵ Bilan d'activité de la fédération UNA Nord (2018)

décembre 2018. Elle représentait **24.69%** des clients des structures adhérentes à la fédération UNA Nord. Parmi ces clients, 1404 étaient bénéficiaires de l'APA ce qui veut dire que ces clients avaient un niveau de GIR qui se situait entre 4 (moyennement dépendant) et 1 (très dépendant). En 2018, 27.67⁶% des bénéficiaires des services de la Maison de l'Aide à Domicile touchaient cette allocation.

4. Le modèle économique

La partie revue de la littérature et la partie contexte permettent de dégager un nombre important d'hypothèses sur les éléments qui peuvent influencer nos variables d'intérêts. L'étude cherche à estimer l'impact social des services d'aide à domicile d'UNA Nord et de la Maison de l'Aide à Domicile de Lille pendant la crise sanitaire de la Covid-19. Pour cela, les variables d'intérêts sont les décès et les hospitalisations. En effet, le but est de mettre en lumière le rôle des services d'aide à domicile de la Maison de l'Aide à Domicile et des structures adhérentes à la fédération UNA Nord pendant la période de confinement.

a) Les caractéristiques sociodémographiques

Hypothèse 1 : L'âge devrait entraîner une hausse des hospitalisations et une augmentation du nombre de décès

L'étude porte sur une population particulière correspondant aux bénéficiaires de l'aide à domicile. Cette population se caractérise par une majorité de personnes âgées. Ces personnes sont plus fragiles et plus touchées par les maladies ou les chutes. Elles ont donc plus de chance d'être hospitalisées ou de décéder.

Hypothèse 2 : Le fait d'être une femme devrait augmenter le nombre de décès et d'hospitalisations

Cela est propre à la population étudiée. En effet, les bénéficiaires de l'aide à domicile sont en majorité des femmes comme le montre l'article de Farajallah, M. & al. (2014). En effet, elles ont plus souvent recours aux services d'aide à domicile car elles vivent souvent seules. En effet, les femmes ont une espérance de vie qui est plus élevée que celle des hommes. Elles ont donc une probabilité plus importante d'être seules. Cela explique qu'elles aient plus recours aux services d'aide à domicile. Ainsi, dans les services d'aide à domicile, il y a plus de femmes qui décèdent ou qui sont hospitalisées.

Hypothèse 3 : Les personnes âgées qui sont mariées ont moins de chance de décéder ou d'être hospitalisées dans les services d'aide à domicile

L'hypothèse 2 indiquait que les femmes composaient la majorité des individus qui ont recours à l'aide à domicile car elles se retrouvaient seules après la mort de leurs conjoints. Dans ce cas, il semble logique de supposer que les personnes qui

⁶Rapport d'activité de la MAD (2018)

bénéficient le plus des services d'aide à domicile soient des personnes célibataires ou veuves. De plus, la solitude peut entraîner des problèmes qui peuvent renforcer la dépendance. Ainsi, le fait d'être seul devrait augmenter le nombre de décès.

Hypothèse 4 : Le fait de vivre dans une commune riche ou dans une commune très peuplée devrait diminuer le nombre de décès.

En effet, les personnes qui vivent dans des communes avec un revenu médian élevé devraient avoir plus facilement accès financièrement aux ressources nécessaires pour vieillir en bonne santé. D'un autre côté, une personne qui vit dans une ville très peuplée aura plus facilement accès aux services de première nécessité comme les magasins, les pharmacies ou les centres de soins médicaux. Ainsi, les personnes qui vivent dans ces deux catégories de communes devraient subir moins de décès que les personnes qui vivent dans des communes isolées, c'est-à-dire peu peuplées et/ou pauvres.

b) L'état de santé des bénéficiaires de l'aide à domicile

Hypothèse 5 : Une dégradation de l'état de santé devrait entraîner une augmentation du nombre de décès et d'hospitalisations.

Avec l'âge les individus sont plus fragiles. Ces derniers sont souvent touchés par des maladies chroniques qui les obligent à aller régulièrement à l'hôpital. Les services d'aide à domicile prennent en charge majoritairement des personnes âgées dépendantes. En effet, une personne âgée qui a de gros problèmes physiques ou psychiques et qui ne peut plus se déplacer aura recours à un service pour faire ses courses, pour faire sa toilette, pour faire son ménage et pour préparer ses repas, etc. Le niveau de dépendance est mesuré par le GIR. Il est possible d'avoir un biais de sélection négatif. En effet, si on part de l'éventualité que les personnes les plus dépendantes qui font appel au service d'aide à domicile ont un risque plus important d'aller à l'hôpital ou de mourir. Une personne moins dépendante n'aurait pas recours à l'aide à domicile et un risque moindre d'être hospitalisée ou de décéder. On peut alors sous-estimer l'effet de l'aide à domicile.

Hypothèse 6 : La période de l'année influence grandement le nombre de décès et le nombre d'hospitalisations. Ainsi, en hiver et en été, le nombre de décès devrait augmenter.

La saison peut fortement influencer le nombre de décès. En effet, on peut remarquer que le nombre de décès est plus important en hiver. En effet, cette période de l'année est marquée par une abondance d'épidémies comme celle de la grippe. Celles-ci entraînent généralement une augmentation des décès. Le nombre de décès est aussi élevé durant l'été. En effet, l'été est souvent marqué par des phases de canicule qui entraînent un nombre de décès important pour les personnes âgées, qui sont parfois isolées. Ces 2 saisons enregistrent un nombre de décès et d'hospitalisations plus important que l'automne.

Hypothèse 7 : Le but de cette étude est d'étudier l'impact des services d'aide à domicile sur le nombre de décès et d'hospitalisations pendant la phase de confinement qu'a connu la France entre mars et mai 2020. On peut supposer que le nombre de décès n'a pas augmenté dans les services d'aide à domicile.

D'après l'INSEE, le nombre de décès devrait augmenter pendant la période de confinement. En effet, on remarque qu'il existe une surmortalité dans le département du Nord pendant la période de confinement. Celle-ci est plus forte par rapport à la même période en 2019 et en 2018. On peut donc s'attendre à observer un signe positif du confinement sur le nombre de décès. Du côté des hospitalisations, on peut s'attendre à un effet négatif du confinement. En effet, d'un côté, on a pu voir précédemment que le nombre d'hospitalisations a augmenté pour les patients atteints du coronavirus. Mais, d'un autre côté, le nombre d'hospitalisations pour d'autres causes que le coronavirus connaît une baisse de moitié dans l'hôpital d'Arras, par exemple. Donc, on peut s'attendre à une baisse du nombre d'hospitalisations pendant la période de confinement puisque le nombre de cas de coronavirus était de 10.9 cas pour 100 000 habitants dans le département du Nord. Ce département enregistre 3254 hospitalisations. Donc, le nombre d'hospitalisations devrait baisser pendant le confinement.

IV. Les données

Les deux sections précédentes ont permis de définir les hypothèses qui vont être testées dans cette étude et d'éclairer le contexte de celle-ci. L'enjeu de cette partie est de décrire les données qui vont être utilisées pour répondre à la question.

1. Les données mobilisées pour cette étude

Pour effectuer cette étude des données ont été mobilisées. Il s'agit des données récupérées auprès des structures qui appartiennent au réseau (données collectées via une enquête par questionnaire) d'une part et d'autre part des données de l'INSEE.

a) L'enquête effectuée auprès des structures

L'enquête a été effectuée auprès de toutes les structures de service d'aide et d'accompagnement à domicile qui sont membres du réseau UNA Nord. Cela correspond à 17 structures qui se répartissent sur l'ensemble du département du Nord. Cette enquête, tout comme l'étude était constituée de deux volets. Le premier volet correspond aux décès et le deuxième volet correspond aux hospitalisations. Au niveau des décès, chaque structure devait renseigner la **date du décès de l'individu** bénéficiaire. Ainsi, nous disposons de la date complète avec le jour, le mois et l'année de décès de l'individu. Cette date de décès est comprise entre le 1^{er} janvier 2019 et le 31 mai 2020 pour maximiser le nombre de réponses tout en ayant un maximum de données pour évaluer l'impact. L'âge des individus décédés a également été demandé. En effet, on peut supposer qu'avec le vieillissement des individus, le nombre de décès augmente. Le niveau de GIR a aussi été collecté auprès de chaque individu. Ce niveau de GIR correspond à un niveau de

dépendance. Celui-ci varie entre 1 et 6. Le niveau 6 correspond à une personne âgée indépendante. A l'inverse, le niveau 1 correspond à une personne âgée très dépendante. Ce niveau de GIR peut donner accès à des aides du département. En effet, à partir du GIR 4 les individus ont accès à l'allocation personnalisée d'autonomie qui est une aide pour les personnes âgées dépendantes versées par les départements. Ensuite, le genre des individus décédés a été demandé aux structures. Pour finir, la commune de chaque individu a été demandée.

Au niveau des hospitalisations, on retrouve à peu près les mêmes variables. En effet, il a été demandé aux structures de fournir l'âge, le genre, le niveau de dépendance et la commune de chaque individu hospitalisé. En plus de cela, **la date d'entrée et la date de sortie de l'hôpital** ont été demandées. Ainsi, la base de données est composée de tous les individus qui sont entrés à l'hôpital entre le 1^{er} janvier 2019 et le 31 mai 2020.

Sur 17 structures avec un service d'aide et d'accompagnement à domicile adhérentes à la fédération UNA Nord seulement **10 ont répondu**. De plus, parmi ces 10 structures certaines n'ont pas fourni les informations nécessaires pour être intégrées à la base de données car ces dernières ne portaient que sur la période de confinement. Celles-ci ont été supprimées pour éviter une surestimation du nombre de décès ou d'hospitalisations pendant le confinement. Un autre problème apparu dans les réponses des structures est la présence de données agrégées. Ainsi, une structure a fourni le nombre total de décès en 2019 et en 2020. Ces données n'ont donc pas pu être exploitées et cette structure n'a pas été intégrée à la base de données. Au final, pour les décès, on retient 8 structures répondantes. Pour les hospitalisations, on peut retenir 7 structures répondantes. Les structures conservées ont au moins une observation avant le confinement et une observation pendant la période de confinement. Cette enquête a été menée sur une période très courte qui s'étend du **13 juin au 30 juin 2020**. (*Voir annexe 1 : enquête auprès des structures*)

b) Une deuxième source de données : l'INSEE

La deuxième source de données utilisée dans cette étude provient de l'INSEE. En effet, pour étudier la surmortalité dans le département du Nord, le fichier détaillé des décès a été exploité. Cette base de données était constituée de l'année, du mois et du jour de décès. De plus, on retrouve dans cette base de données le département de décès et la commune de décès. Pour chaque individu décédé on retrouve également son année, son mois et son jour de naissance.

Deux autres bases de données de l'INSEE ont été exploitées pour catégoriser les communes. En effet, dans la partie précédente l'une des hypothèses mentionne des catégories de communes. Pour construire ces catégories, il fallait récupérer le **revenu** par commune et la **population** par commune. Pour récupérer le niveau de revenu par commune, le fichier FiLoSoFi (2019) a été exploité. Ce fichier comprend l'ensemble des communes de France au 1^{er} janvier 2017. De plus, on peut trouver le revenu disponible médian par unité de consommation en euros. Pour finir, le fichier Populations légales 2017 (INSEE, 2019) a été exploité pour récupérer les données de la population totale pour chaque commune du Nord en 2017.

Les sources	L'enquête	L'INSEE	L'auteur
Les catégories de données			
Les individus	-L'âge des individus -Le genre de l'individu -Le GIR de l'individu -La commune de résidence de l'individu <i>Pour la MAD</i> -La situation familiale de certains individus -La catégorie socioprofessionnelle des individus		-L'âge des individus de la MAD calculé à partir des dates de naissance et de décès des individus ou le 31 mai 2020.
Le temps	-La date de décès -La date de début d'hospitalisations -La date de fin d'hospitalisations		-la saison -Le confinement -Le nombre de décès et d'hospitalisations
Les communes		-Le revenu médian des ménages de la commune -La population de la commune	-3 catégories pour les communes selon le revenu médian -4 catégories selon la population

Tableau 4 : Tableau récapitulatif des différentes sources de données. La colonne auteur correspond aux variables qui ont été construites par l'auteur de l'étude à partir des données provenant des 2 autres sources.

2. Les variables utilisées pour les modèles

a) Les variables utilisées pour les décès

La première difficulté rencontrée pour faire cette étude consistait à récupérer les données des structures. La MAD est la première structure à avoir répondu. Cette base de données contenait plusieurs variables intéressantes pour notre analyse. En effet, dans cette dernière, il était possible de retrouver l'ensemble des individus décédés sur une période assez longue. En effet, cet échantillon était constitué de 937 individus décédés sur une période allant de janvier 2017 à mai 2020. Cette base de données était composée de l'âge des individus qui correspond à la différence entre la date de décès et la date de naissance de l'individu. Elle contenait également le niveau de dépendance des individus. Celui-ci est une variable intéressante puisqu'elle donne une indication sur l'état de santé du client. Néanmoins, le niveau de GIR ne sera pas intégré dans la suite de l'analyse car, cette information n'est disponible que pour certains individus. En effet, toutes les personnes âgées ne sont pas systématiquement évaluées. D'autres informations comme le statut marital ou la catégorie socioprofessionnelle étaient disponibles dans cette première base. Mais, celles-ci n'étaient pas disponibles pour l'ensemble des individus décédés. Elles n'ont donc pas été exploitées. De plus, ces variables n'étaient disponibles que pour la maison de l'aide à domicile.

En plus de ces variables, le genre des individus était également disponible. La commune de l'individu est présente et donne une information très importante. En effet, elle permet de connaître les possibilités données à un individu d'accéder aux soins ou à des denrées alimentaires. A ces informations, nous avons ajouté le nom de la structure. Cette base de données a été réduite à la période du 1^{er} janvier 2019 au 31 mai 2020 qui correspond aux périodes demandées aux autres structures. Les informations des autres structures ont été ajoutées au fur et à mesure de leur réception.

Au total, notre échantillon de décès contient 620 personnes décédées. Jusqu'à maintenant nous avons les informations de base sur les individus décédés. Néanmoins, il n'était pas encore possible de tester la totalité des hypothèses formulées à la fin de la partie 1. En effet, pour cela, il manquait la dimension temporelle et des informations plus précises sur les communes. Le revenu médian et la population ont donc été ajoutés pour chaque commune dans notre base de données. La clé de fusion des tables était la commune. A partir de ces deux informations, les communes ont été séparées en plusieurs catégories : 4 pour la population et 3 pour le revenu médian. Ainsi, en dessous de 5000 habitants, il y a les villages. Entre 5 000 et 20 000 habitants, on retrouve les petites villes. Entre 20 000 et 100 000 habitants, on trouve les villes. Pour finir, au-delà de 100 000 habitants, il y a les grandes villes. Au niveau du revenu médian, on retrouve les villes pauvres qui ont un niveau de revenu médian inférieur à 19 500€. Ensuite, il y a les communes moyennes qui ont un niveau de revenu compris entre 19 500 et 25 700€. Pour finir, les communes riches ont un revenu médian supérieur à 25 700€.

La dimension temporelle s'exprime à travers la date de décès de tous les individus. Elle correspond au mois et à l'année de décès de chaque individu. A partir de cette information, 3 nouvelles variables ont été créées. Une variable muette qui correspond à la période de confinement. Celle-ci est définie comme la période qui s'étend de mars à mai 2020, une variable saison qui correspond aux quatre saisons de l'année et une variable qui correspond au nombre de décès à chaque mois. Dans cette base de données on retrouve également l'âge et le nombre de décès en logarithme. Cela permettra de faciliter l'interprétation des estimations. (Annexe : *Tableau 9*)

Pour résumer notre base de décès correspond à un empilement mensuel d'échantillon aléatoire de personnes décédées sur une période de 17 mois. L'échantillon final se compose **de 620 individus et de 27 variables**.

b) Les variables utilisées pour les hospitalisations

Pour les hospitalisations, la structure de la base de données est à peu près la même. Pour cette base de données, les structures ont fourni leurs données pour la même période de temps. Ainsi, on retrouve des variables comme l'âge, le niveau de dépendance, le genre et la commune de résidence de chaque individu qui est entré à l'hôpital entre le 1^{er} janvier 2019 et le 31 mai 2020. De nouveau, chaque commune a ensuite été regroupée en catégorie.

Ces catégories dépendent du niveau de population dans chaque commune et du revenu médian pour ces communes. Les catégories sont les mêmes que pour celle des décès. Ainsi, au niveau du revenu, on retrouve trois catégories de communes qui sont les communes riches, les communes pauvres et les communes moyennes. Au niveau de la population, on retrouve les quatre catégories de communes avec les villages, les petites villes, les villes et les grandes villes.

Pour finir, on retrouve aussi la dimension temporelle. En effet, on utilise la date d'entrée à l'hôpital pour définir les variations temporelles. Ainsi, on récupère le mois et l'année d'entrée à l'hôpital. A partir de cette information 3 nouvelles variables ont été créées. Une variable muette qui correspond à la période de confinement. Celle-ci est définie comme la période qui s'étend de mars à mai 2020, une variable saison qui correspond aux quatre saisons de l'année et une variable qui correspond au nombre d'hospitalisations à chaque mois. La date de sortie d'hospitalisations est également fournie par les structures. Cela nous permet de calculer le temps d'hospitalisation de chaque individu en jour. Dans cette base de données on retrouve également l'âge, le nombre d'hospitalisations et les temps d'hospitalisations en logarithme. Cela permettra de faciliter l'interprétation des estimations. (Annexe : *Tableau 10*)

Pour résumer, notre base d'hospitalisations correspond à un empilement mensuel d'échantillon aléatoire de personnes hospitalisées sur une période de 17 mois. L'échantillon final se compose de **2808 individus hospitalisés et de 29 variables**. Il est à noter que dans cet échantillon, on peut retrouver le même individu plusieurs fois.

3. L'analyse descriptive

Cette partie a pour objectif de présenter quelques statistiques descriptives sur les deux échantillons. Cette analyse vise dans un premier temps à décrire l'échantillon global. Ensuite, nous comparerons l'échantillon avant et après le confinement et une analyse graphique basée sur les variables à expliquer. Pour finir, les corrélations entre les variables seront étudiées.

a) L'analyse descriptive sur l'échantillon des décès statistiques descriptives

Variables	Obs.	Moyenne	Ecart type	Min.	Max.
age	618	80.65	12.83	21.00	106.00
genre	620	0.58	0.49	0	1
Confinement	620	0.20	0.40	0	1
nb.deces	620	39.10	10.31	24	58
communes_riches	620	0.03	0.18	0	1
communes_moyennes	620	0.45	0.50	0	1
communes_pauvres	620	0.52	0.50	0	1
villages	620	0.21	0.41	0	1

petites_villes	620	0.33	0.47	0	1
villes	620	0.27	0.45	0	1
grandes_villes	620	0.19	0.40	0	1
automne	620	0.13	0.34	0	1
ete	620	0.15	0.35	0	1
printemps	620	0.37	0.48	0	1
hiver	620	0.35	0.48	0	1

Tableau 5 : Statistiques descriptives de l'échantillon de décès

Dans un premier temps on peut remarquer que la base de données des décès sur les structures adhérentes à la fédération UNA Nord est composée de 620 individus. Sur ces individus on ne connaît pas l'âge de 2 individus. On peut également constater que la majorité des personnes dans cet échantillon sont des personnes âgées. En effet, **l'âge moyen** dans cet échantillon est de **80.65 ans** à plus ou moins 12 ans. Comme évoqué précédemment les femmes bénéficient plus souvent des services d'aide à domicile. En effet, on peut observer dans les associations de la fédération UNA Nord que **58%** des personnes décédées sont des **femmes**.

Seulement **20%** des individus sont morts pendant la **période de confinement**. Celle-ci correspond aux mois de **mars à mai 2020**. La plupart des individus décédés vivaient dans des communes pauvres. En effet, 52% des individus décédés vivaient dans une commune avec un revenu médian inférieure à 19 500€. On remarque aussi que seulement 3% des individus sont morts dans des communes avec un revenu médian supérieur à 25 700€. 33% des individus sont morts dans les villes qui ont une population qui se situe entre 5 000 et 20 000 habitants. Les autres villes sont à peu près réparties de façon égale entre les quatre catégories. En effet, chaque catégorie représente entre 19 et 33% des personnes décédées. Les saisons ont aussi une influence sur le nombre de décès. En effet, il est possible d'observer que le nombre de décès est plus important pendant le printemps et l'hiver soient 37% et 35%. On remarque que c'est pendant l'automne que le nombre de décès est le plus faible.

En moyenne, **le nombre de décès** sur la période qui s'étend de janvier 2019 à mai 2020 est **de 39.10**. Ces premières statistiques descriptives permettent de mettre en exergue certaines hypothèses.

Pour continuer cette partie, il semble intéressant d'étudier l'évolution de ces statistiques entre la période précédant le confinement et la période de confinement (annexe : *Tableau 11*). En moyenne, les personnes qui sont décédées avant et pendant le confinement avaient à peu près le même âge. Ainsi, il n'y a pas de différence significative entre la période d'avant confinement et la période de confinement. Au niveau du genre, on retrouve toujours une majorité de femmes qui sont décédées dans l'échantillon. Il n'y a pas de différence significative au niveau du genre des individus décédés avant et pendant le confinement.

Parmi les individus qui vivaient dans une commune riche, il n'y a pas eu de différence significative entre les deux périodes des décès avec un risque d'erreur de 5%. Pour les deux catégories de villes suivantes, il n'y a pas de différence significative entre les deux périodes dans l'échantillon. Si, on observe les catégories de ville construite selon la population, alors il n'y a pas de différence significative entre les décès qui ont eu lieu avant et pendant le confinement.

Dans le même temps, il semble important d'étudier également les corrélations entre les variables. Ainsi, le genre est très peu corrélé avec les autres variables explicatives, à part l'âge. Il existe une **corrélation positive entre l'âge et le fait d'être une femme**. La variable liée au **confinement est positivement corrélée au nombre de décès**. On observe également une corrélation négative entre le confinement et les différentes saisons. Seul le printemps est corrélé positivement à la variable confinement. Cela s'explique par le fait que la variable confinement correspond aux mois du printemps. On peut aussi voir une corrélation du nombre de décès avec les saisons. Ainsi, le nombre de décès est négativement corrélé à l'automne et l'été. A l'inverse, il existe une corrélation positive entre le nombre de décès, le printemps et l'hiver. Ainsi, pendant ces deux saisons le nombre de décès augmente (annexe : *Tableau 12*).

Pour finir, cette analyse descriptive sur l'échantillon des décès, Il semble intéressant d'étudier graphiquement l'évolution des décès sur la période d'étude. On obtient alors le graphique suivant :

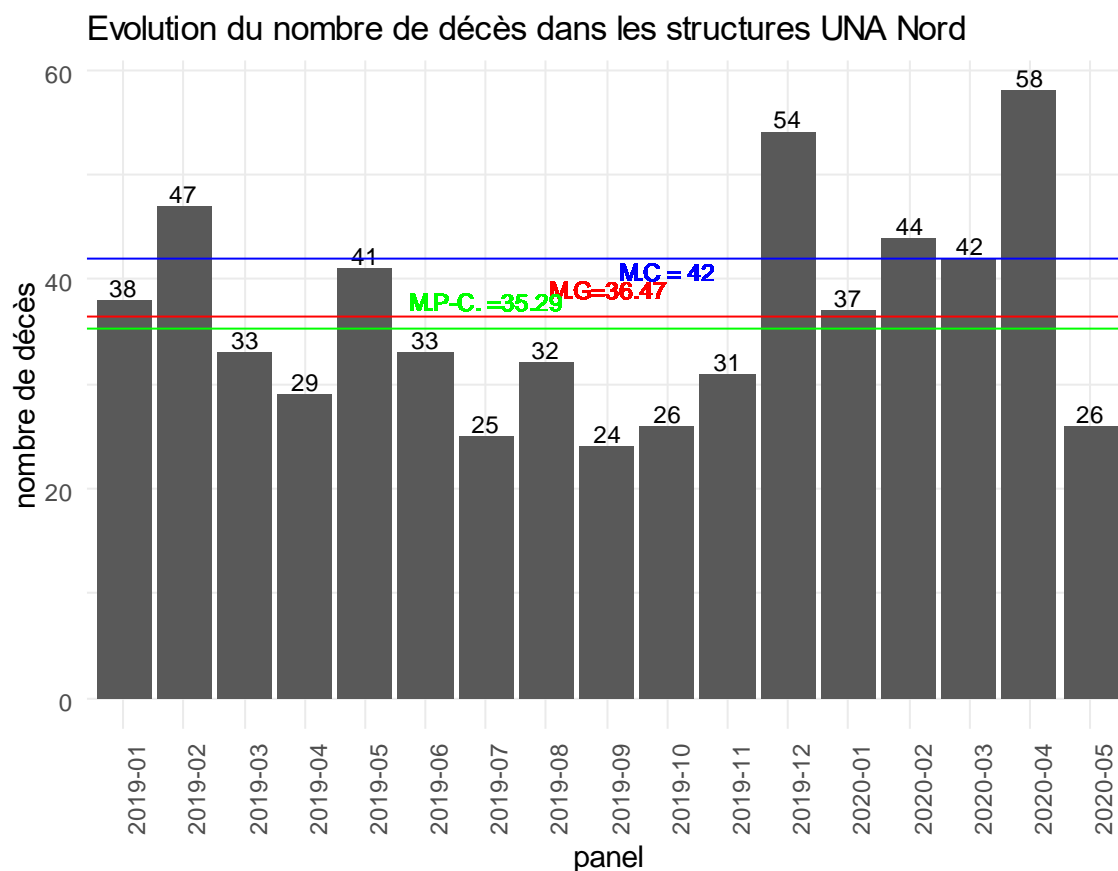


Figure 4 : Evolution et moyennes mensuelles des décès. M.G représente la moyenne mensuelle sur la totalité de la période. M.C est la moyenne mensuelle pendant le confinement. M.P-C représente la moyenne mensuelle avant le confinement.

On peut voir sur ce graphique l'évolution des décès dans les différentes structures du réseau UNA. On peut remarquer que le nombre de décès le plus important apparaît au mois d'avril 2020. Ce mois-ci le nombre de décès atteignait 58. On peut aussi observer que le nombre de décès est plus faible pendant l'automne. En moyenne, avant le confinement, le nombre de décès était de 35.29. Durant le confinement, le nombre de décès moyen était de 42. On a donc une **augmentation du nombre de décès de 6.71**. Il semble important d'effectuer un test de comparaison des moyennes entre la moyenne avant et pendant le confinement. On cherche à tester si le nombre de décès moyen pendant le confinement est significativement différent du nombre de décès avant le confinement. Ce test aboutit à la conclusion **que le nombre de décès n'est pas significativement différent** entre la période qui précède le confinement et la période de confinement.

b) L'analyse descriptive sur l'échantillon des hospitalisations

Statistiques descriptives

Variables	Obs.	Moyenne	Ecart-type	Min.	Max.
age	2,780	79.87	12.69	15.00	105.00

tps_hospi	2,808	137.69	1486.51	1	29,583
confinement	2,808	0.14	0.34	0	1
genre	2,808	0.67	0.47	0	1
nb_hospitalisation	2,808	169.78	28.12	105	236
communes_riches	2,808	0.02	0.16	0	1
communes_moyennes	2,808	0.49	0.50	0	1
communes_pauvres	2,808	0.49	0.50	0	1
villages	2,808	0.20	0.40	0	1
petites_villes	2,808	0.35	0.48	0	1
villes	2,808	0.24	0.43	0	1
grandes_villes	2,808	0.21	0.41	0	1
automne	2,808	0.17	0.37	0	1
ete	2,808	0.18	0.38	0	1
printemps	2,808	0.33	0.47	0	1
hiver	2,808	0.33	0.47	0	1

Tableau 6 : Statistiques descriptives sur l'échantillon des individus hospitalisés

Dans un premier temps on peut remarquer que la base de données des hospitalisations sur les structures adhérentes à la fédération UNA Nord est composée de 2808 individus. Sur ces individus, on ne connaît pas l'âge de 28 personnes. On peut également constater que la majorité des individus dans cet échantillon sont des personnes âgées. En effet, l'âge moyen dans cet échantillon est de 79.87 ans à plus ou moins 12 ans. Comme évoqué précédemment les femmes bénéficient plus souvent des services d'aide à domicile. En effet, on peut observer dans les associations de la fédération UNA Nord que 67% des personnes hospitalisées sont des femmes.

Seulement **14%** des individus ont été **hospitalisés pendant la période de confinement**. Celle-ci correspond aux mois de mars à mai 2020. La plupart des individus hospitalisés vivaient dans des communes moyennes. En effet, 49% des individus hospitalisés vivaient dans une commune avec un revenu médian compris entre 19 500€ et 25 700€. On remarque aussi que seulement 2% des individus sont hospitalisés dans des communes avec un revenu médian supérieur à 25 700€. 35% des individus sont hospitalisés dans les villes qui ont une population qui se situe entre 5 000 et 20 000 habitants. Les autres villes sont à peu près réparties de façon égale entre les quatre catégories. En effet, chaque catégorie représente entre 20 et 35% des personnes hospitalisées. Les saisons ont aussi une influence sur le nombre d'hospitalisations. En effet, il est possible d'observer que le nombre de décès est plus important pendant le printemps et l'hiver soit 33%. On remarque que c'est pendant l'automne que le nombre d'hospitalisations est le plus faible. Le temps moyen d'hospitalisations dans l'échantillon est de 137.69 jours.

En moyenne, **le nombre d'hospitalisations** sur la période qui s'étend de janvier 2019 à mai 2020 est en **moyenne de 169.78**. Ces premières statistiques descriptives permettent de mettre en exergue certaines hypothèses.

Pour continuer cette partie, il semble intéressant d'étudier l'évolution de ces statistiques entre la période précédant le confinement et la période de confinement (annexe : Tableau 13). En moyenne, les personnes qui sont hospitalisées avant et pendant le confinement avaient à peu près le même âge. Ainsi, il n'y a pas de différence significative entre la période d'avant confinement et la période de confinement. Au niveau du genre, on retrouve toujours une majorité de femmes qui sont décédées dans l'échantillon. Il n'y a pas de différence significative au niveau du genre des individus hospitalisés avant et pendant le confinement. Le temps d'hospitalisations est en moyenne plus long pendant le confinement. Néanmoins, la différence n'est pas significative entre les deux périodes.

Parmi les individus qui vivaient dans une commune riche, il n'y a pas eu de différence significative entre les deux périodes des hospitalisations. Pour les deux catégories de villes suivantes, il n'y a pas de différence significative entre les deux périodes dans l'échantillon. Si, on observe les catégories de ville construites selon la population, alors il n'y a pas de différence significative entre les hospitalisations qui ont eu lieu avant et pendant le confinement.

Dans le même temps, il semble important d'étudier également les corrélations entre les variables. Ainsi, le genre est très peu corrélé avec les autres variables explicatives, à part l'âge. En effet, il existe une corrélation positive entre l'âge et le fait d'être une femme. **La variable liée au confinement est négativement corrélée au nombre d'hospitalisations.** On observe également une corrélation négative entre le confinement et les différentes saisons. Seul le printemps est corrélé positivement à la variable confinement. Cela s'explique par le fait que la variable confinement correspond aux mois du printemps. Le temps d'hospitalisations est peu corrélé avec les autres variables. On peut aussi voir une corrélation du nombre d'hospitalisations avec les saisons. Ainsi, le nombre d'hospitalisations est négativement corrélé à l'automne, l'été et le printemps. Cela signifie que durant ces trois saisons le nombre d'hospitalisations baisse. A l'inverse, il existe une corrélation positive entre le nombre d'hospitalisations et l'hiver. Ainsi, pendant cette saison le nombre d'hospitalisations augmente (annexe : *Tableau 14*).

Pour finir cette analyse descriptive sur l'échantillon des hospitalisations, il semble intéressant d'étudier graphiquement l'évolution des entrées à l'hôpital sur la période d'étude. On obtient alors le graphique suivant :

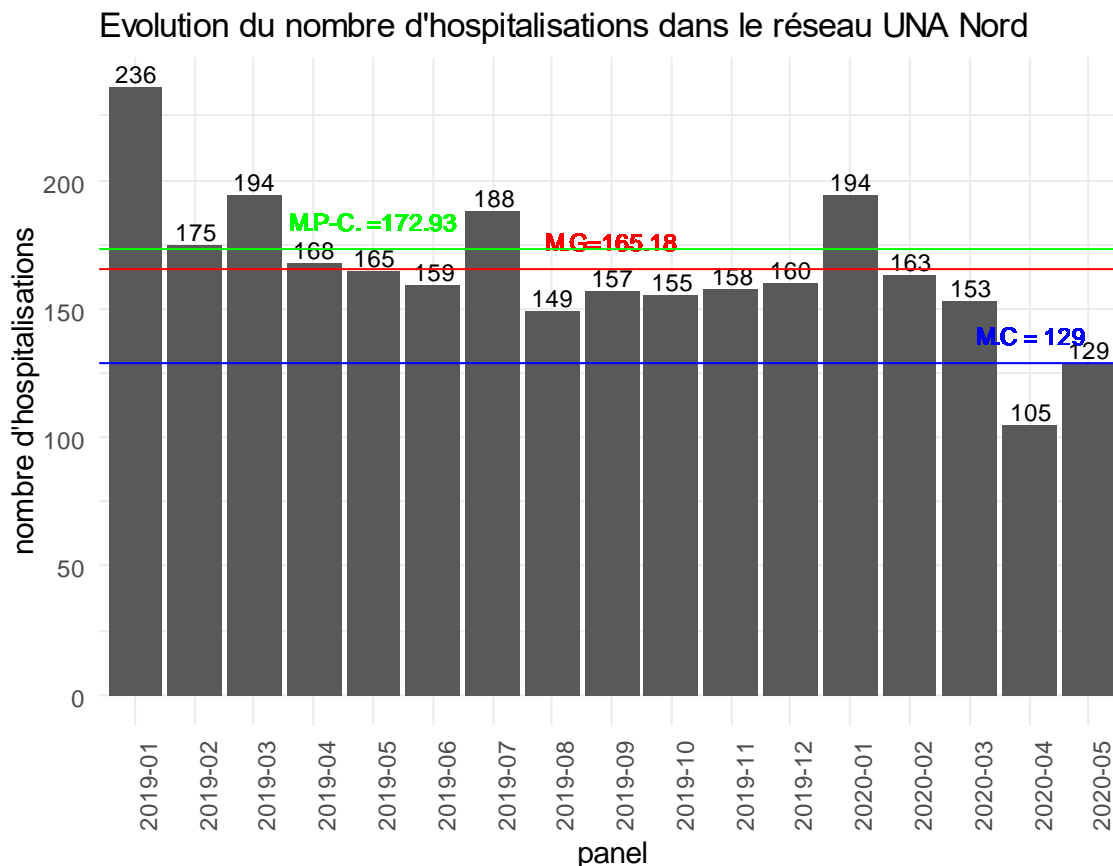


Figure 5: évolution des moyennes mensuelles d'hospitalisations dans les 7 structures du réseau UNA Nord. M.G représente la moyenne mensuelle sur la totalité de la période. M.C est la moyenne mensuelle pendant le confinement. M.P-C représente la moyenne mensuelle avant le confinement.

On peut voir sur ce graphique l'évolution des hospitalisations dans les différentes structures du réseau UNA. On peut remarquer que le nombre d'entrées à l'hôpital le plus important apparaît en hiver. En effet, le nombre maximum d'hospitalisations est obtenu au mois de janvier 2019, durant lequel le nombre d'hospitalisations atteignait 236. On peut aussi observer que le nombre d'hospitalisations est plus faible pendant l'automne. En moyenne, avant le confinement, le nombre d'hospitalisations était de 172.93. Durant le confinement, le nombre d'hospitalisations moyen était de 129. **On a donc une baisse du nombre de d'entrée de 43.93.** Il semble important d'effectuer un test de comparaison des moyennes entre la moyenne avant et pendant le confinement. On cherche à tester si le nombre d'entrées moyen pendant le confinement est significativement inférieur au nombre d'hospitalisations avant le confinement. Ce test aboutit à la conclusion que le nombre d'entrées à l'hôpital a **significativement baissé** entre la période qui précède le confinement et la période de confinement.

V. La méthodologie d'estimation

Cette partie cherche à expliquer la méthodologie économétrique qui est utilisée pour estimer les paramètres de nos modèles pour les hospitalisations et pour les décès dans les structures du réseau UNA Nord.

1. La méthodologie d'estimation

Les données mobilisées pour cette étude ont une forme particulière. Celle-ci ne correspond pas aux méthodes classiques d'estimation. En effet, les données ne correspondent ni à une coupe transversale pure dans laquelle nous n'observerions que des individus sans variations temporelles, ni à une pure série temporelle où nous n'observerions que des variations temporelles. Les données utilisées ici combinent ces deux dimensions. Ainsi, les données sont des données en coupe transversales agrégées à travers le temps ou **Pooled Cross Sections across Time** (Leblond, S. 2003). Chaque individu observé dans notre échantillon de personnes décédées n'apparaît qu'une seule fois. On ne suit donc pas le même groupe d'individu à travers le temps. Il n'est donc pas possible d'utiliser les méthodes de panel. En effet, les données utilisées ici sont des données en coupe transversale agrégée indépendante. Il s'agit, en effet, d'échantillon aléatoire constitué chaque mois entre janvier 2019 et mai 2020. Le fait d'agréger des échantillons aléatoires à différents points dans le temps permet d'augmenter la taille de l'échantillon. Cela permet d'obtenir des estimateurs plus précis et des tests statistiques plus robustes (Woodridge, J. 2009). Pour faire apparaître les variations temporelles, il suffit de créer des variables muettes pour les périodes que l'on souhaite étudier. Ainsi, dans cette étude qui vise à mesurer l'impact des services d'aide à domicile pendant le confinement, 4 variables muettes ont été créées. Elles servent à capturer l'effet temporel du confinement et des saisons. L'automne est choisi comme base et est intégré à la constante du modèle. L'automne a été choisi car il s'agit d'une saison pour laquelle aucune hypothèse n'a été formulée. De plus, il s'agit de la saison avec le nombre de décès et d'hospitalisations le moins élevé.

Ce type de données permet de construire un modèle de régression multiple qui sera estimé par la méthode des moindres carrés ordinaires. Dans ce type de régression, une variable dépendante Y est expliquée par un certain nombre de variables explicatives notées X . Cela permet d'obtenir les valeurs des coefficients des variables indépendantes X . Autrement, dit le modèle peut s'écrire :

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

Où Y est la variable dépendante que l'on cherche à expliquer. β représente l'ensemble des coefficients à estimer et X représente l'ensemble des variables indépendantes qui expliquent Y . u est l'erreur de mesure du modèle. Les coefficients de ces modèles sont estimés par la méthode des moindres carrés ordinaires. Cette méthode cherche à minimiser la somme des carrés des erreurs. Ainsi, cette méthode d'estimation cherche à minimiser u .

Ce type de modèle repose sur cinq hypothèses qui permettent d'obtenir des estimations robustes et convergentes :

- L'espérance mathématique de l'erreur est nulle
- La variance de l'erreur est constante pour toutes les observations
- La matrice des X est non-aléatoire
- Le modèle est correctement spécifié
- La matrice X est de plein rang

Sous ces 5 hypothèses les estimateurs sont sans biais et à variance minimale. Ainsi, il s'agit des meilleurs estimateurs linéaires sans biais.

Néanmoins, avec les données disponibles l'hypothèse 2 a de grandes chances de ne pas être respectée. En effet, comme expliquée plus haut, les données disponibles sont des données en coupe transversale agrégée dans le temps. Ainsi, les données ont deux dimensions : une dimension individuelle et une dimension temporelle. Or, d'un côté, on dispose de données individuelles qui ont de grande chance de souffrir d'hétéroscédasticité c'est-à-dire que les variances varient pour chaque individu. De l'autre côté, les données temporelles ont de grandes chances de souffrir d'autocorrélation. Il y a donc très peu de chance pour que les estimateurs soient à variance minimale.

Pour faire face à cette difficulté, une autre méthode d'estimation est nécessaire. Celle-ci doit permettre de revenir à des estimateurs à variance minimale. Or, un estimateur qui possède les mêmes propriétés que l'estimateur des moindres carrés ordinaires mais qui permet de corriger le problème d'hétéroscédasticité et d'autocorrélation est l'estimateur des moindres carrés généralisés. Cet estimateur va chercher à corriger la matrice des variances covariances. En effet, dans le cas où la variance n'est pas constante pour toutes les observations, il faut introduire un nouvel élément qui est la matrice des variances covariances corrigée. On obtient ainsi, l'estimateur de β qui devient:

$$\hat{\beta} = (X' \Omega^{-1} X)^{-1} (X' \Omega^{-1} Y) \quad (2)$$

On obtient aussi la matrice des variances covariances suivantes :

$$V(\hat{\beta}) = \sigma^2_u (X' \Omega^{-1} X)^{-1} \quad (3)$$

L'estimateur des moindres carrés généralisés s'écrit de la façon suivante

$$\hat{\Omega}^{-1/2} Y = \hat{\Omega}^{-1/2} X\beta + \hat{\Omega}^{-1/2} u \quad (4)$$

L'enjeu de cette méthode d'estimation est d'estimer la matrice des variances covariances des résidus (Bourbonnais, R. 2015).

Finalement, la méthode d'estimation des modèles sur les décès et les hospitalisations sera la méthode des moindres carrés ordinaires. Cette méthode sera corrigée grâce à la méthode des moindres carrés généralisés si des problèmes d'hétéroscédasticité ou d'autocorrélation sont détectés.

2. La construction des modèles économétriques

a) Le modèle économétrique sur les décès

Après avoir présenté les méthodes d'estimation qui seront utilisées pour mesurer l'impact des services d'aide à domicile de la fédération UNA Nord pendant le confinement, il semble pertinent de justifier la construction des modèles économétriques choisis.

Ainsi, pour expliquer le nombre de décès au sein des structures adhérentes au réseau UNA Nord, Il semble intéressant de conserver l'âge des individus. En effet, celui-ci donne une première idée de l'état de santé de l'individu. Ainsi, les personnes plus âgées devraient avoir une probabilité plus importante de décéder. En effet, les personnes âgées sont plus touchées par des maladies chroniques (Mayo-Simbsler, S. & Viquant, JP, 2010). Ces personnes sont plus fragiles. De plus, l'âge augmente aussi la probabilité de chute ce qui accroît la possibilité d'un décès. Donc, l'âge possède un effet positif sur le nombre de décès. Ensuite, le genre des individus est intégré au modèle. En effet, cette variable impact le nombre de décès. Comme on l'a vu précédemment, l'échantillon est composé d'une majorité de femmes. Celles-ci vivent plus longtemps que les hommes et leur âge avancé augmente le risque de décès. Cette variable devrait avoir un effet positif sur le nombre de décès.

Les caractéristiques des communes ont aussi un rôle important. En effet, il y a trois variables qui correspondent aux caractéristiques des villes selon le revenu. Il s'agit de variable muette. Deux de ces variables sont introduites dans le modèle. Ainsi, les communes riches et les communes pauvres sont ajoutées au modèle. Les communes moyennes sont intégrées à la constante pour que le modèle puisse être estimé. Ainsi, le fait de vivre dans une commune riche devrait avoir un effet négatif sur le nombre de décès. Ainsi, le fait de vivre dans une commune riche devrait entraîner une baisse du nombre de décès. Des variables catégorielles ont aussi été introduites pour contrôler l'effet de la population sur les communes. Ainsi, on introduit 3 variables catégorielles pour contrôler le niveau de population. On conserve les villages, les petites villes et les villes. Les grandes villes sont intégrées à la constante. Ainsi, le nombre de décès devrait baisser dans les grandes villes.

Pour contrôler l'effet des saisons sur le nombre de décès, on introduit trois variables indicatrices. Ainsi, en hiver et en été le nombre de décès devrait augmenter. Ainsi, l'effet de l'hiver et de l'été devrait être positif sur le nombre de décès. Pour finir, la variable la plus importante qui est intégrée à notre modèle et qui permet de répondre à notre question est la variable indicatrice qui correspond au confinement. On peut supposer que le nombre de décès n'augmentera pas dans les services d'aide à domicile du réseau UNA Nord. Ainsi, le modèle peut s'écrire de la façon suivante :

$$\log_deces_t = \beta_0 + \beta_1 \log_age_i + \beta_2 genre_i + \beta_3 communes_riche_i + \beta_4 communes_pauvres_i + \beta_5 villages_i + \beta_6 petites_villes_i + \beta_7 villes_i + \beta_8 printemps_t + \beta_9 hiver_t + \beta_{10} ete_t + \beta_{11} \mathbf{confinement}_t + u_{it} \quad (5)$$

Dans ce modèle, le logarithme du nombre de décès varie dans le temps. En effet, le nombre de décès correspond au nombre d'individus qui sont décédés chaque mois. β_0 représente la constante du modèle. β_1 représente l'effet de l'âge des individus sur le nombre de décès. Cet estimateur varie en fonction des individus. β_2 représente l'effet d'être une femme sur le nombre de décès. Cet estimateur varie selon le sexe de l'individu. β_3 représente l'effet de l'individu qui vit dans une commune riche sur le nombre de décès. Cet estimateur varie en fonction de la commune de l'individu. β_4 montre l'effet de l'individu qui vit dans une commune pauvre sur le nombre de décès. Cet indicateur varie en fonction de la commune de l'individu. β_5 représente l'effet de l'individu qui vit dans un village sur le nombre de

décès. Cet estimateur varie en fonction de la commune de l'individu. β_6 représente l'effet de l'individu qui vit dans une petite ville sur le nombre de décès. Cet effet dépend de la commune de l'individu. β_7 montre l'effet de l'individu qui vit dans une ville sur le nombre de décès. Cet effet varie selon la commune de l'individu. β_8 représente l'effet de décéder au printemps sur le nombre de décès. Cet effet varie dans le temps. β_9 est l'effet de décéder pendant l'hiver sur le nombre de décès. Cet effet varie dans le temps. β_{10} montre l'effet de décéder en été sur le nombre de décès. Cet effet varie dans le temps. **Pour finir, β_{11} représente l'effet du confinement sur le nombre de décès. Cet effet varie dans le temps.** U_{it} représente l'erreur de mesure qui varie selon les individus et selon le temps.

Pour ce modèle, une analyse en log-niveau a été utilisée. En effet, celle-ci possède certains avantages. Premièrement elle permet d'estimer des semi-élasticités. Elle permet donc une interprétation en pourcentage qui est plus simple que l'interprétation en valeur absolue. De plus, le passage à l'échelle logarithmique devrait permettre dans une certaine mesure de supprimer l'effet taille et donc de lutter contre l'hétéroscédasticité. Avec ce modèle, on ne peut pas estimer un des élasticités, dans le sens où la plupart des variables sont des indicatrices qui prennent soit la valeur 1 soit la valeur 0. Or, le logarithme de zéro n'existe pas, c'est pourquoi un modèle log-log n'aurait pas de sens.

b) Le modèle économétrique sur les hospitalisations

La modélisation sur les hospitalisations est très proche de celles sur les décès. En effet, de façon générale les individus qui vont à l'hôpital ont des chances de décès plus importantes que les personnes qui n'y vont pas. Ainsi, on peut s'attendre à ce que les individus qui ont été hospitalisés meurent. Néanmoins, cette hypothèse n'est pas vérifiable dans les échantillons puisque les individus ne sont pas identifiés. Néanmoins, il est possible de penser que les mêmes facteurs influencent les hospitalisations et les décès. Ainsi, on retiendra l'âge qui devrait avoir un effet positif sur le nombre d'hospitalisations. Le genre devrait aussi avoir un impact positif sur le nombre d'hospitalisations.

On observe les mêmes caractéristiques des communes. Ainsi, on peut supposer que le fait de vivre dans une commune avec un revenu médian élevé possède un effet négatif sur le nombre d'entrées à l'hôpital. A l'inverse, le fait de vivre dans une commune avec un revenu médian faible devrait avoir un effet positif sur le nombre d'hospitalisations. Au niveau de la population, il est envisageable de supposer que les individus qui vivent dans des villes ou des grandes villes aient un effet positif sur le nombre d'entrée à l'hôpital.

Le temps passé à l'hôpital est ajouté au modèle. Cet indicateur peut donner une idée sur la gravité de la maladie d'une personne. Donc, le temps d'hospitalisation devrait avoir un impact positif sur le nombre d'hospitalisations. Ensuite, il existe aussi un effet saisonnier. Celui-ci est contrôlé par l'introduction des saisons dans les variables explicatives. On peut s'attendre aux mêmes effets. Ainsi, l'hiver et l'été devraient avoir un impact positif sur le nombre d'hospitalisations. Pour finir, le confinement devrait avoir un effet négatif sur le nombre d'hospitalisations. En effet, le confinement a eu plusieurs conséquences. Comme on a pu le voir dans la

partie III sur le contexte de l'étude, le nombre d'hospitalisations hors coronavirus a fortement baissé. Certains individus redoutant la contamination ne sont pas allés à l'hôpital. Ainsi, la presse a pu relayer une forte baisse de la fréquentation de certains hôpitaux comme à Arras ou à Paris peu importe les causes (la voix du nord, 2020), (Godeluck, S. 2020). Le modèle peut donc s'écrire de la façon suivante :

$$\text{Log_hospi}_t = \beta_0 + \beta_1 \log_age_i + \beta_2 \log_tps_t + \beta_3 \text{genre}_i + \beta_4 \text{communes_riche}_i + \beta_5 \text{communes_pauvres}_i + \beta_6 \text{villages}_i + \beta_7 \text{petites_villes}_i + \beta_8 \text{villes}_i + \beta_9 \text{printemps}_t + \beta_{10} \text{ete}_t + \beta_{11} \text{hiver}_t + \beta_{12} \text{confinement}_t + u_{it} \quad (6)$$

Dans ce modèle, le logarithme des hospitalisations varie en fonction du temps. En effet, le nombre d'hospitalisations correspond au nombre d'entrées à l'hôpital qui ont eu lieu chaque mois. β_0 représente la constante du modèle c'est-à-dire le nombre d'hospitalisations qui ont eu lieu sans aucune autres variables explicatives. β_1 représente l'effet de l'âge sur le nombre de décès. Il s'agit d'une élasticité. L'âge varie en fonction des individus. β_2 correspond à l'effet du temps moyen de séjour à l'hôpital sur le nombre d'hospitalisations. Il s'agit d'une élasticité. Le temps d'hospitalisations varie en fonction du temps. β_3 représente l'effet d'être une femme sur le nombre d'hospitalisations. Cette variable varie selon les individus. Il s'agit d'une semi-élasticité. β_4 correspond à l'impact de vivre dans une commune avec un revenu médian supérieur à 25 700€ sur le nombre d'hospitalisations. Cette variable varie en fonction de la commune habitée par l'individu. Il s'agit d'une semi-élasticité. β_5 correspond à l'effet de vivre dans une commune avec un revenu médian inférieur à 19 500€ sur le nombre d'hospitalisations. Cette variable varie en fonction de la commune de résidence de l'individu. Ces deux variables s'interprètent par rapport au fait de vivre dans une commune moyenne. β_6 représente l'effet de vivre dans un villages par rapport à une grande ville sur le nombre d'hospitalisations. β_7 mesure l'effet de vivre dans une petite ville par rapport à une grande ville sur le nombre d'hospitalisations. Il s'agit d'une semi-élasticité comme toutes les variables qui vont suivre. Cette variable varie en fonction de commune de résidence de l'individu. β_8 correspond à l'effet de vivre dans une ville par rapport à une grande ville sur le nombre d'hospitalisations. Cette variable varie en fonction de la commune de l'individu. β_9 représente l'effet du printemps sur le nombre d'hospitalisations par rapport à l'automne. Cette variable varie en fonction du mois d'entrée à l'hôpital. β_{10} représente l'effet de l'été par rapport à l'automne sur le nombre d'entrées à l'hôpital. β_{11} correspond à l'effet de l'hiver sur le nombre d'hospitalisations par rapport à l'automne. **Pour finir, β_{12} représente l'effet du confinement sur le nombre d'hospitalisations.**

Dans ce modèle, on utilise deux types de spécification. Une spécification log-log pour les variables quantitatives. Cette spécification permet de lutter contre l'effet taille de l'échantillon. Cela pourrait permettre de régler le problème de l'hétéroscédasticité qui pourrait apparaître avec des données en coupe transversale. Cette spécification permet aussi de faciliter l'interprétation puisque les coefficients s'interprètent comme des élasticités. Autrement dit, ils s'interprètent sous forme de pourcentages. Pour les variables indicatrices on ne peut pas les transformer en log à car elles prennent la valeur 1 en cas de survenu de l'événement et 0 sinon. Or le logarithme de zéro n'existe pas. Donc, ces estimateurs seront interprétés comme des

semi-élasticités. Cela permet aussi d'avoir une interprétation sous forme de pourcentage.

VI. Les résultats

Après avoir présenté la méthodologie d'estimation et les modèles qui seront estimés, nous présenterons dans cette partie l'analyse des résultats. Celle-ci sera construite autour de trois parties. Ces parties sont :

- l'analyse des résultats sur les décès
- l'analyse des résultats sur les hospitalisations
- l'étude des hospitalisations évitées et des économies générées

1. L'analyse des décès

La partie précédente a permis de sélectionner les variables qui vont être introduites dans le modèle. Ce modèle vise à expliquer le logarithme du nombre de décès en fonction de plusieurs variables. Le modèle qui a été retenu pour expliquer le logarithme du nombre de décès correspond à l'équation (5) de la partie précédente. Ce modèle est dans un premier temps estimé par la méthode des moindres carrés ordinaires. Il sera ensuite estimé par la méthode des moindres carrés généralisés. Avec ces 2 méthodes d'estimations, on obtient les résultats suivants :

résultat d'estimation sur les décès			
	<i>Variable dépendante :</i>		
	Log_deces		
	MCO	MCG	MCG corrigé
log_age	-0.0540 (0.0408)	-0.0540 (0.0408)	-0.0374 (0.0348)
genre	0.0278* (0.0151)	0.0278* (0.0151)	0.0161 (0.0123)
communes_riches	-0.0672 (0.0430)	-0.0672 (0.0430)	-0.0558 (0.0357)
communes_pauvres	-0.0110 (0.0181)	-0.0110 (0.0181)	-0.0183 (0.0161)
villages	-0.0442* (0.0265)	-0.0442* (0.0265)	-0.0376 (0.0254)
petites_villes	-0.0350	-0.0350	-0.0373*

	(0.0246)	(0.0246)	(0.0222)
viles	-0.0206	-0.0206	-0.0389**
	(0.0229)	(0.0229)	(0.0195)
hiver	0.4932***	0.4932***	0.4985***
	(0.0238)	(0.0238)	(0.0192)
printemps	0.2458***	0.2458***	0.2346***
	(0.0273)	(0.0273)	(0.0199)
ete	0.1052***	0.1052***	0.1042***
	(0.0282)	(0.0282)	(0.0184)
Confinement	0.2424***	0.2424***	0.2687***
	(0.0243)	(0.0243)	(0.0258)
Constant	3.5550***	3.4994***	3.4950***
	(0.1782)	(0.1782)	(0.1512)
Observations	618	618	618
R ²	0.5346		
Adjusted R ²	0.5261		
Log Likelihood		145.5622	206.9542
Akaike Inf. Crit.		-265.1244	-383.9084
F Statistic	63.2805*** (df = 11; 606)		
Note:	() : Ecart-type, * p<0.1; ** p<0.05; *** p<0.01		

Tableau 7 : Résultats d'estimation par les moindres carrés ordinaires et par les moindres carrés généralisés

a) Analyse de la robustesse des estimations par les MCO

Cette partie a pour but de justifier l'usage des moindres carrés généralisés. A première vue, l'estimation par les moindres carrés ordinaires est plutôt bonne. En effet, le modèle qui est construit ici explique **53.46%** de la variance totale du logarithme des décès. Cela s'explique notamment par le fait que l'on travaille sur de l'humain. Donc, beaucoup d'autres variables explicatives pourraient expliquer cette variance. Ensuite, le résultat du test de Fischer montre que le modèle est **globalement significatif**. Il n'y a pas de problème de multi-colinéarité entre les variables. En effet, le Tableau 19 (annexe2) montre que le facteur d'inflation de la variance (VIF) est inférieur à 4 pour toutes les variables.

Néanmoins, ce modèle risque de donner des résultats biaisés. En effet, l'analyse des résidus et du qqplot normal (Annexe : *Figure 6*) montre bien que les résidus ont peu de chance de suivre une loi normale. En effet, les résidus sont assez éloignés de la droite théorique. Pour vérifier l'hypothèse de normalité des résidus, le test de Wilk-Shapiro est utilisé (Annexe : *Tableau 16*). Il ressort que les résidus ne suivent pas une loi normale. L'idée serait maintenant de vérifier l'hypothèse 2, c'est-à-dire de vérifier la présence d'autocorrélation et d'hétéroscédasticité des résidus. Le test de Breusch-Pagan permet de détecter la présence d'hétéroscédasticité dans les résidus (Annexe : *Tableau 17*). Les coefficients estimés par la méthode des moindres

carrés ordinaires ne sont pas à variance minimale. Il est possible de détecter la présence d'autocorrélation des résidus. Pour tester cette possibilité, le test de Durbin-Watson va être utilisé (Annexe : *Tableau 18*). Ce test retourne une statistique de test de 1.2908. Les résidus sont donc positivement corrélés entre eux. De plus, l'étude de l'autocorrélogramme des résidus montre la présence d'une autocorrélation significative des résidus entre eux (Annexe : *Figure 9*). Les estimations par les moindres carrés ordinaires font face aux deux problèmes possibles qui font que les estimateurs ne sont pas à variance minimale. Cela a une incidence sur la significativité des paramètres estimés.

Pour corriger ce problème, le modèle va être estimé par la méthode des moindres carrés généralisés. Cette méthode permet de transformer le modèle pour obtenir des estimateurs à variance minimale. Pour cela, une structure autorégressive d'ordre 1 est ajoutée afin de corriger l'autocorrélation des erreurs. Pour corriger la présence d'hétéroscédasticité une pondération a été ajoutée pour corriger la matrice des variance-covariance (Annexe : *Figure 8*). Le modèle MCG corrigé (AIC = -383.91) est plus puissant que le modèle MCG sans correction (AIC = -265.12).

b) L'interprétation des résultats obtenus par les MCG

La partie précédente permet de justifier l'usage des moindres carrés généralisés. Seuls les coefficients estimés par les moindres carrés généralisés vont être interprétés car ceux-ci sont plus justes que les estimateurs des moindres carrés ordinaires. Les décès baissent de 3.73% dans les petites villes, par rapport aux grandes villes de plus de 100 000 habitants. Ensuite, le fait d'habiter dans une ville entraîne une baisse du nombre de décès de 3.89%.

En hiver, le nombre de personnes décédées augmente de 46.39% par rapport à l'automne. Cela est conforme aux hypothèses formulées dans la partie III.4. En effet, il semble logique que le nombre de décès augmente en hiver à cause de l'augmentation des épidémies à cette époque de l'année. Au printemps, le nombre de décès augmente de 24.01% par rapport à l'automne. Durant l'été, le nombre de décès augmente de 11.02% par rapport à l'automne. Cela correspond à l'intuition. En effet, comme expliqué précédemment l'été est une saison qui est marquée par des périodes de forte chaleur. Ces périodes peuvent avoir un effet important sur le nombre de décès.

La période de confinement est marquée par une augmentation des décès de 26.87% par rapport à la période qui précède. Ce résultat est contradictoire avec ce qui était attendu. En effet, on pouvait supposer que le nombre de décès aurait baissé pendant la période de confinement dans les services de la fédération UNA Nord.

Un facteur qui pourrait expliquer ce phénomène est l'année de référence utilisée (2019). Dans la partie III, un graphique montre l'évolution du nombre de décès pendant la période de confinement. On peut voir que l'année 2019 est l'année où il y a le moins de décès. Cela peut donc expliquer l'augmentation du nombre de décès. Pour vérifier, cette hypothèse il faudrait disposer de données qui remonteraient à une période antérieure à 2019. Cela est possible à partir des

données d'une structure adhérente à la fédération UNA Nord : la MAD. **On peut voir que si l'on ajoute l'année 2018, alors le confinement possède toujours un impact significatif** (Annexe : *Tableau 20*). Pour résumer, avec l'échantillon disponible, les services d'aide à domicile de la fédération UNA Nord n'ont pas réussi à limiter le nombre de décès pendant le confinement. En effet, ce nombre de décès a augmenté pendant la période de confinement qui s'étale sur le mois de mars avril et mai 2020 par rapport à la période qui précède.

A partir des données de la MAD il est aussi possible d'effectuer des comparaisons spatiales entre les décès qui ont eu lieu au sein de la structure et la totalité des décès qui ont eu lieu dans le département du Nord ou dans la région des Hauts-de France. Ces comparaisons montrent que le confinement a un impact plus important au sein de la maison de l'aide à domicile (**29.16%**) que dans le département du Nord (**5.91%**) ou dans la région des Hauts-de-France (**10.71%**). (Annexe : *Tableau 21*)

Néanmoins, les résultats des estimations pour le département et la région sont biaisés par l'autocorrélation des résidus. Cela peut être une explication des différences observées. Une autre explication serait que les populations sont trop différentes. Ainsi, les trois populations auraient un comportement différent vis-à-vis des facteurs explicatifs du confinement. D'après ce tableau, des caractéristiques comme l'âge, le genre, et le nombre de décès sont significativement différents. (Annexe : *Tableau 22*)

Il est important de préciser que ces estimations ne sont valables que sur l'échantillon étudié et ne peuvent pas être extrapolées au niveau de la France, à cause du biais de sélection. En effet, les données disponibles ne permettent pas d'être sûr que l'échantillon soit représentatif des services d'aide à domicile. Pour cela, il faudrait des données sur les autres structures et sur les autres fédérations. Cela permettrait d'obtenir un échantillon représentatif des services d'aide à domicile.

2. L'analyse des hospitalisations

La partie précédente a permis de sélectionner les variables qui vont être introduites dans le modèle. Ce modèle vise à expliquer le logarithme du nombre d'hospitalisations en fonction de plusieurs variables. Le modèle qui a été retenu pour expliquer le logarithme du nombre d'hospitalisations correspond à l'équation (6) de la partie précédente. Ce modèle est dans un premier temps estimé par la méthode des moindres carrés ordinaires. Il sera ensuite estimé par la méthode des moindres carrés généralisés. Avec ces 2 méthodes d'estimations, on obtient les résultats suivants :

résultat d'estimation sur les hospitalisations			
Variable dépendante:			
Log_hospi			
MCO	MCG	MCG corrigé	correction de white
(1)	(2)	(3)	(4)

log_age	0.0285** (0.0120)	0.0285** (0.0120)	0.0283** (0.0120)	0.0285** (0.0112)
genre	-0.0034 (0.0048)	-0.0034 (0.0048)	-0.0035 (0.0048)	-0.0034 (0.0048)
communes_riches	-0.0054 (0.0144)	-0.0054 (0.0144)	-0.0053 (0.0145)	-0.0054 (0.0132)
communes_pauvres	-0.0021 (0.0056)	-0.0021 (0.0056)	-0.0021 (0.0056)	-0.0021 (0.0054)
villages	-0.0068 (0.0080)	-0.0068 (0.0080)	-0.0070 (0.0080)	-0.0068 (0.0077)
petites_villes	-0.0058 (0.0075)	-0.0058 (0.0075)	-0.0060 (0.0075)	-0.0058 (0.0074)
villes	-0.0010 (0.0069)	-0.0010 (0.0069)	-0.0013 (0.0069)	-0.0010 (0.0071)
ete	0.0584*** (0.0075)	0.0584*** (0.0075)	0.0584*** (0.0074)	0.0584*** (0.0046)
hiver	0.1790*** (0.0066)	0.1790*** (0.0066)	0.1790*** (0.0066)	0.1790*** (0.0049)
printemps	0.1173*** (0.0074)	0.1173*** (0.0074)	0.1173*** (0.0074)	0.1173*** (0.0033)
Log_tps	0.0017 (0.0013)	0.0017 (0.0013)	0.0016 (0.0013)	0.0017 (0.0013)
confinement	-0.2997*** (0.0078)	-0.2997*** (0.0078)	-0.2997*** (0.0077)	-0.2997*** (0.0085)
Constant	4.9329*** (0.0521)	4.9329*** (0.0521)	4.9329*** (0.0520)	4.9329*** (0.0485)
Observations	2,780	2,780	2,780	
R ²	0.5129			
Adjusted R ²	0.5107			
Log Likelihood		2,008.32	2,009.08	
Akaike Inf. Crit.		-3,988.63	-3,988.17	
F Statistic	242.76*** (df = 12; 2767)			

Note: () : Ecart-type, * p<0.1; ** p<0.05; *** p<0.01

Tableau 8 : Résultats d'estimation sur les hospitalisations par les moindres carrés ordinaires et par les moindres carrés généralisés entre janvier 2019 et mai 2020

a) Analyse de la robustesse des estimations par les MCO

Cette partie a pour but de justifier l'usage des moindres carrés généralisés. A première vue, l'estimation par les moindres carrés ordinaires est plutôt bonne. En effet, le modèle qui est construit ici explique **51.29%** de la variance totale du logarithme des hospitalisations. Cela s'explique notamment par le fait que l'on travaille sur de l'humain. Donc, beaucoup d'autres variables explicatives pourraient expliquer cette variance. Ensuite, le résultat du test de Fischer montre que le **modèle est globalement significatif**. Il n'y a pas de problème de multi-colinéarité entre les variables. En effet, le Tableau 23 (annexe) montre que le facteur d'inflation de la variance (VIF) est inférieur à 4 pour toutes les variables (Annexe : *Tableau 23*).

Néanmoins, ce modèle risque de donner des résultats biaisés. En effet, l'analyse des résidus et du qqplot normal (Annexe : *Figure 10*) montre bien que les résidus ont peu de chance de suivre une loi normale. En effet, les résidus sont assez éloignés de la droite théorique. Pour vérifier l'hypothèse de normalité des résidus, le test de Wilk-Shapiro est utilisé (Annexe : *Tableau 25*). Il ressort que les résidus ne suivent pas une loi normale. L'idée serait maintenant de vérifier l'hypothèse 2, c'est-à-dire de vérifier la présence d'autocorrélation et d'hétéroscédasticité des résidus. Le test de Breusch-Pagan permet de détecter la présence d'hétéroscédasticité dans les résidus (Annexe : *Tableau 26*). Les coefficients estimés par la méthode des moindres carrés ordinaires ne sont pas à variance minimale. Il est possible de détecter la présence d'autocorrélation des résidus. Pour tester cette possibilité, le test de Durbin-Watson va être utilisé (*Tableau 27*). Ce test retourne une statistique de test de 1.91. Les résidus ne sont donc pas corrélés entre eux. Les estimations par les moindres carrés ordinaires font face au problème d'hétéroscédasticité. Ce problème fait que les estimateurs ne sont pas à variance minimale. Cela a une incidence sur la significativité des paramètres estimés.

Pour corriger ce problème, le modèle va être estimé par la méthode des moindres carrés généralisés. Cette méthode permet de transformer le modèle pour obtenir des estimateurs à variance minimale. Une correction par la méthode de White est également possible. On peut également observer que le modèle estimé par les moindres carrés généralisés avec une pondération pour corriger l'hétéroscédasticité n'est pas plus performant (AIC=-3988.17) que le modèle sans correction (AIC = -3988.63) car il ne permet pas de minimiser l'AIC

b) L'interprétation des résultats par les MCG et par la correction de White

L'estimation par les moindres carrés généralisés et par la correction de White donne des résultats similaires pour les coefficients. Ainsi, dans les deux cas le fait de vieillir entraîne une augmentation du nombre d'hospitalisations. Ainsi, une augmentation de l'âge de l'individu de 1% entraîne une augmentation du nombre d'hospitalisations de 0.03%. Ce résultat correspond aux hypothèses. En effet, on pouvait s'attendre à ce que l'âge ait un impact positif sur le nombre d'hospitalisations. Il semble logique qu'une personne qui vieillit soit touchée par un nombre plus important de maladies, dont des maladies chroniques. Celles-ci risquent de l'obliger à aller plus souvent à l'hôpital.

En été, les bénéficiaires de l'aide à domicile ont tendance à se rendre plus souvent à l'hôpital qu'en automne. En effet, durant l'été, le nombre d'hospitalisations augmente de 5.84% par rapport à l'automne. Cela semble cohérent avec les intuitions. En effet, durant l'été des épisodes de fortes chaleurs ont lieu. Ces épisodes peuvent entraîner des déshydratations importantes. Cela peut conduire à une augmentation du nombre d'hospitalisations. Pendant, l'hiver le nombre d'hospitalisations augmente également par rapport à l'automne. Ainsi, durant l'hiver le nombre d'hospitalisations augmente de 17.90% par rapport à l'automne. Ceci est cohérent puisque cette saison est marquée par une forte recrudescence des épidémies. Au printemps, le nombre d'hospitalisations augmente de 11.73% par rapport à l'automne.

Pour finir, durant la période de confinement (mars, avril et mai 2020), le nombre d'hospitalisations a fortement baissé. En effet, **le nombre d'entrées à l'hôpital pendant le confinement a baissé de 29.97%** par rapport à la période précédente, dans les structures adhérentes à UNA Nord. Il n'est pas possible de conclure que c'est grâce à l'intervention des services d'aide à domicile que le nombre d'hospitalisations a baissé. En effet, on ne peut pas être certain de la présence d'un lien causal entre ces deux phénomènes. En effet, dans la partie III on a remarqué que le nombre de séjour aux urgences avait fortement baissé pour des causes différentes du coronavirus. **Ainsi, il est seulement possible d'affirmer que le nombre d'hospitalisations a baissé dans les structures d'aide à domicile de la fédération UNA Nord pendant le confinement.**

Il est important de préciser que ces estimations ne sont valables que sur l'échantillon étudié et ne peuvent pas être extrapolées au niveau de la France, à cause du biais de sélection. En effet, les données disponibles ne permettent pas d'être sûr que l'échantillon soit représentatif des services d'aide à domicile. Pour cela, il faudrait des données sur les autres structures et sur les autres fédérations. Cela permettrait d'obtenir un échantillon représentatif des services d'aide à domicile.

3. L'estimation des économies générées pendant la période de confinement

L'estimation précédente a permis de mettre en valeur la baisse des hospitalisations pendant le confinement, dans les structures adhérentes aux réseaux UNA Nord. Or, le coût d'une prise en charge à l'hôpital est très élevé quel que soit le service de prise en charge. Néanmoins, il existe de fortes disparités entre les services. En effet, selon une étude de l'ATIH (Agence Technique de l'Information sur l'Hospitalisation) portant sur les coûts de prise en charge à l'hôpital, on peut remarquer de très fortes disparités dans les coûts moyens. Cette disparité se retrouve entre les établissements publics et privés et entre les secteurs d'activités. En 2012, dans les hôpitaux publics, le coût moyen d'une prise en charge en chirurgie était de 4 564€, contre 1 910€ dans les cliniques privées. En médecine générale, le coût moyen d'une prise en charge à l'hôpital public était de 2620€ contre 1554€ dans une clinique privée (ATIH, 2012).

Dans le cadre de cette étude, les coûts moyens en médecine ont été retenus. En effet, utiliser le coût moyen total n'aurait pas d'intérêt dans le sens où celui-ci

prend en compte des activités qui ne concernent pas les bénéficiaires de l'aide à domicile qui constituent l'échantillon. C'est le cas par exemple des prises en charge en néonatalogie, en obstétrique, etc. On prendra en compte le coût moyen en médecine car celui-ci représente la majorité des séjours en 2012. Pour calculer, le coût moyen d'une prise en charge à l'hôpital, une moyenne pondérée a été calculée entre les hôpitaux privés et publics en France. Cet indicateur a été calculé de la façon suivante :

$$\frac{nb \text{ séjour méd pu} * \text{coût moyen méd pu} + nb \text{ séjour méd privé} * \text{coût moyen méd privé}}{nb \text{ séjour méd}} \quad (7)$$

Où nb séjour méd pu représente le nombre de séjours nationaux en médecine dans les hôpitaux publics en 2012, coût moyen méd pu représente le coût moyen d'une prise en charge en médecine dans un hôpital public en 2012, nb séjour méd privé représente le nombre de séjours nationaux en médecine dans un hôpital privé en 2012, le coût moyen méd privé correspond au coût moyen d'une prise en charge en médecine dans un hôpital privé, en 2012. Pour finir, nb séjour méd représente le nombre de séjours total en médecine. On aboutit à un **coût moyen pour une prise en charge à l'hôpital, en médecine de 2 496.18€**

Dans les modèles précédents une spécification log-log ou log-niveau a été utilisée. Celle-ci permettait de faciliter l'interprétation de chaque coefficient puisque celle-ci était donnée sous forme relative. Cette spécification pose problème maintenant puisque l'on cherche le nombre d'hospitalisations qui ont pu être évitées pendant le confinement. Cela nous permettra d'estimer les économies liées à la baisse des hospitalisations pendant le confinement, en termes de prise en charge à l'hôpital.

Les résultats d'estimation font ressortir les mêmes effets que précédemment (*Tableau 288*). Durant le confinement, **44.66 bénéficiaires des services d'aide à domicile de la fédération UNA Nord ne sont pas allés à l'hôpital par rapport à la période précédente**. La période de confinement a été définie comme une période de 3 mois qui va de mars à mai 2020. Cela correspond à une durée de 92 jours. Afin de chiffrer l'économie que représente la baisse des visites à l'hôpital pendant cette période, on effectue le calcul suivant :

$$\begin{aligned} & \text{Nombre d'hospitalisations évitées} * \text{durée du confinement} \\ & * \text{le coût moyen en médecine} \end{aligned} \quad (8)$$

On obtient, ainsi, $44.66 * 2496.18 * 92 = 10\,256\,104.69\text{€}$. On constate que pendant le confinement, le nombre d'hospitalisations évitées dans les structures d'aide à domicile du **réseau UNA Nord a permis de générer plus de 10 millions d'euros d'économie par rapport à la période précédente**.

VII. Conclusion

Cette étude a cherché à montrer l'efficacité des services d'aide à domicile pendant le confinement lié à la crise sanitaire du coronavirus.

Elle s'est concentrée sur l'efficacité des structures adhérentes au réseau UNA Nord. Celle-ci a été mesurée par le biais des décès et des entrées à l'hôpital qui peuvent être expliqués par de nombreux facteurs. Parmi ces facteurs, il y a l'âge, le genre, la commune de résidence, la saison, le temps d'hospitalisation et la période de confinement.

L'ensemble de ces données a été fourni par les structures qui sont adhérentes aux réseaux UNA Nord. Certaines des variables explicatives varient au cours du temps. C'est le cas des saisons ou du confinement. D'autres variables varient en fonction des individus. C'est le cas de l'âge ou du genre. Ainsi, la base de données regroupe les 2 dimensions. On parle dans notre cas de coupe transversale agrégée à travers le temps. Notre dimension temporelle correspond à chaque mois entre janvier 2019 et mai 2020. Ce type de série peut faire face à deux problèmes. Dans un premier temps, avec des données en coupe transversale les erreurs risquent d'être hétéroscédastiques. Ensuite, la présence d'une variabilité temporelle sur 17 périodes risque de faire apparaître un problème d'autocorrélation. Ces deux problèmes vont entraîner un biais dans les estimations. En effet, certains coefficients ne seront pas significatifs alors qu'ils devraient l'être à cause du gonflement des variances.

Pour corriger ces problèmes et ainsi, obtenir les meilleurs estimateurs sans biais la méthode des moindres carrés généralisés a été utilisée. Dans cette étude on utilisera la méthode des moindres carrés ordinaires pour estimer les effets des variables sur le nombre de décès et d'hospitalisations. Cet estimateur sera ensuite corrigé par la méthode des moindres carrés généralisés. Les modèles ont été spécifiés en log-log et en log-niveau pour faciliter l'interprétation des résultats obtenus. Pour choisir les modèles, on s'est appuyé sur la revue de la littérature et sur le contexte de l'étude. Ensuite, la qualité des modèles corrigés a été évaluée par le biais du critère d'Akaike (AIC).

La plupart des résultats qui ont été obtenus correspondent aux résultats qui étaient attendus après l'étude de la littérature et du contexte. En effet, ces résultats montrent que les saisons ont bien un impact positif et significatif sur le nombre de décès. De plus, pour chaque saison on obtient bien l'effet attendu. En effet, on pouvait supposer au départ que le nombre de décès et d'hospitalisations en hiver seraient plus importants que pour les autres saisons.

Le seul résultat significatif un peu déroutant correspond à l'effet du confinement sur les décès. On s'attendait à n'obtenir aucun effet. Or, on trouve un effet positif. Mais cela peut s'expliquer par la période de référence qui correspond à l'année 2019. En effet, d'après la figure 1 le nombre de décès en 2019 pendant la période de confinement est plus bas en 2019. On a pu tester sur les données de la maison de l'aide à domicile la présence d'un effet du confinement en partant de

l'année 2018. Dans ce cas le confinement a toujours un impact significatif sur le nombre de décès. Donc, le fait d'ajouter une année ne permet pas d'annuler l'effet significatif du confinement.

Pour les hospitalisations, on retrouve également des résultats qui sont conformes à la littérature et aux intuitions. Ainsi, comme pour les décès, on retrouve l'hiver qui possède le plus fort effet positif sur le nombre d'hospitalisations. On trouve également l'effet de l'âge qui augmente le nombre d'hospitalisations.

Pour terminer on trouve un effet négatif du confinement sur le nombre d'hospitalisations. Ainsi, pendant le confinement, le nombre d'hospitalisations a baissé de 29.97% dans les structures adhérentes à UNA Nord. Cette baisse représente 44.66 hospitalisations évitées. Ainsi, pendant le confinement, le nombre d'hospitalisations a baissé dans les structures par rapport à la période qui précède le confinement. Cette baisse représente plus de 10 millions d'euros d'économie.

Pour aller plus loin dans cette étude, plusieurs points peuvent être soulevés. Le premier point correspond à la disponibilité des données. En effet, l'échantillon n'est composé que des années 2019 et 2020. Cela ne permet pas de montrer l'effet des structures d'aide à domicile pendant le confinement en termes de décès. Il serait donc intéressant d'avoir d'autres années en plus pour conforter ou infirmer nos hypothèses. Un deuxième point important est la présence d'un biais de sélection. En effet, l'échantillon dont nous disposons n'est pas représentatif de la totalité de la population des bénéficiaires de l'aide à domicile ni des structures. Donc, les résultats ne peuvent pas être extrapolés à l'ensemble de la population des bénéficiaires ou des structures. De plus, on ne mesure ici, que la différence dans le temps, on ne corrige donc que le biais temporel. Pour corriger le biais de sélection et ainsi pouvoir réellement mesurer l'impact des services d'aide à domicile, il faudrait comparer les bénéficiaires de l'aide à domicile avec des non bénéficiaires de l'aide à domicile et qu'ils soient statistiquement similaires. Dans ce cas, il serait possible d'utiliser la méthode des doubles différences.

Pour finir, il semble cohérent de penser qu'il existe un biais de variable omise. En effet, cette étude s'appuie sur des individus humains et sur la santé des individus. Or, les données de santé sont des données sensibles qui ne sont pas accessibles.

VIII. Bibliographie

Aouba, A. & al. (2012). Où meurt-on en France ? Analyse des certificats de décès (1993-2008). *Bulletin épidémiologique hebdomadaire*. N°48, p.547-551.

ARS des Hauts-de-France. (2020). *Coronavirus : point de situation en Hauts-de-France*. [<https://www.hauts-de-france.ars.sante.fr/coronavirus-point-de-situation-en-hauts-de-france>]. Consulté le 6 août 2020.

ATIH. (2012). *Les coûts des prises en charge à l'hôpital en médecine, chirurgie et obstétrique*.

[https://www.atih.sante.fr/sites/default/files/public/content/2563/rapport_couts_de_prise_en_charge_2012.pdf]. Consulté le 18 août 2020

Audureau, W. & Vaudano, M. (2020). Coronavirus : du premier cas détecté de Covid-19 au déconfinement, la chronologie d'une crise mondiale. *Le Monde*.

Bretton, L. (2020). Cette crise rend visibles ceux qui sont d'ordinaire invisible. *Libération*

Bourbonnais, R. (2015). Problèmes particuliers : la violation des hypothèses. *Econométrie : Cours et exercices corrigés* (9^e édition, p.125-164). Paris, France : DUNOD

CNSA-DGCS, (2016). *Etude des prestations d'aide et d'accompagnement à domicile et des facteurs explicatifs de leurs coûts*. [https://www.cnsa.fr/documentation/enc_saad_2016_rapport_vdef.pdf]. Consulté le 18 août 2020

Coulibaly, S., Fortin, B. & Isabelle, M. (2020). COVID-19 et maintien à domicile des aînés. *Cirano*. [<https://cirano.qc.ca/files/publications/2020PE-03.pdf>]. Consulté le 18 août 2020.

DARES. (2020). *Les services à la personne en 2018*. [https://www.servicesalapersonne.gouv.fr/files_sap/files/etudes/dares_resultats_services_a_la_personne_2018.pdf]. Consulté le 6 août 2020.

DARES. (2018). *Les services à la personne en 2016*. [<https://dares.travail-emploi.gouv.fr/IMG/pdf/2018-017.pdf>]. Consulté le 6 août 2020

De Brux, J. Goldzahl, L. & Mével, A. (2018). L'aide et les soins à domicile : quelle création de valeurs ? CITIZING pour ADESSADOMICILE et l'OCIRP

Duthil, G. (2007). L'arrivée du privé dans la prise en charge des personnes âgées. *Gérontologie et société*. 30(123), p. 185-200.

Farajallah, M. & Retali, G-R. (2013). Déterminants de l'adoption des services d'aide au maintien à domicile : une évidence empirique. L'exemple de SIGAAL. [https://emnet.univie.ac.at/uploads/media/Farajallah_Retali.pdf]. Consulté le 18 août 2020.

Fontaine, R. & Arnault, L. (2016). *Vers une diminution programmée de l'aide familiale aux personnes âgées en perte d'autonomie ?* [https://www.isere.fr/mda38/Lists/DocumentaryResources/Attachments/6728/GSSweb-aideinformelle_Rapport.pdf]. Consulté le 6 août 2020.

Givord, P. (2014). Méthodes économétriques pour l'évaluation de politiques publiques. *Economie & prévision*. N°204-205, p.1-28

Godeluck, S. (2020). Hors coronavirus, le renoncement aux soins devient très inquiétant. *Les échos*, 18 avril 2020, p.4

Watremetz, L. (2020), *Les masques se retrouvent au cœur de la bataille médico-sociale contre le coronavirus*. [<https://www.hospimedia.fr/actualite/articles/20200323-gestion-des-risques-les-masques-se-retrouvent-au>]. Consulté le 18 août 2020

INSEE. *Tableaux de l'économie française*. 27 février 2020. [<https://www.insee.fr/fr/statistiques/4277754?sommaire=4318291&q=APA>]. Consulté le 7 avril 2020.

INSEE. *France, portrait social*. 19 novembre 2019. [<https://www.insee.fr/fr/statistiques/4238381?sommaire=4238781&q=APA>]. Consulté le 7 avril 2020.

INSEE. *Projections de population à l'horizon 2070*. 3 novembre 2016. [<https://www.insee.fr/fr/statistiques/2496228>]. Consulté le 7 avril 2020.

INSEE. *Evolution du nombre de décès quotidiens et cumulés depuis le 1^{er} mars*. 15 mai 2020. [<https://www.insee.fr/fr/statistiques/4487861?sommaire=4487854>]. Consulté le 17 mai 2020.

INSEE. *Structure et distribution des revenus, inégalité des niveaux de vie en 2016* [Base de données]. 26 juillet 2019. [<https://www.insee.fr/fr/statistiques/4190006>]. Consulté le 7 août 2020.

INSEE. *Téléchargement du fichier d'ensemble des populations légales en 2017* [Base de données]. 30 décembre 2019. [<https://www.insee.fr/fr/statistiques/4265429>]. Consulté le 7 août 2020.

Institut national d'études démographiques. (2020). *La démographie des décès liés à la pandémie COVID-19*.

La Voix du Nord. (2020). *Hôpital d'Arras : le premier mois de crise Covid-19 en chiffres*. 11 juin 2020. [<https://www.lavoixdunord.fr/764075/article/2020-06-11/hopital-d-arras-le-premier-mois-de-crise-covid-19-en-chiffres?referer=%2Farchives%2Frecherche%3Fdatefilter%3Dlastyear%26sort%3Ddate%2Bdesc%26start%3D40%26word%3Dhospitalisation>]

Leblanc-Briot, M-T. (2014). Fin de vie des personnes âgées. *Presses universitaires de Grenoble*. N°119, p. 97-108

Libault, D. (2019), *Concertation Grand âge et autonomie*. [https://solidarites-sante.gouv.fr/IMG/pdf/rapport_grand_age_autonomie.pdf]. Consulté le 26 juin 2020

Mayo-Simbsler, S. & Vinquant, JP. (2010), Vieillissement et prise en charge des maladies chroniques, *Actualité et dossier en santé publique*, n°72, p.33-36

Meslet, E. (2020). Hervé Robert : « Le Covid-19 vient mettre en évidence la fragilité de l'aide à domicile ». *l'Humanité*.

Ministère de l'économie et des finances. (2019). *Les services à la personne –SAP*. [<https://www.servicesalapersonne.gouv.fr/donnees-et-etudes/chiffres-cles>]. Consulté le 7 août 2020.

Rapp, T. & al. (2015). Are public subsidies effective to reduce emergency care ? Emergence from the PLASA study. *Social Science & Medecine*. 138. 31-37 (abstract)

Sands, L.P & al. (2012). Volume of Home and Community Based Services and Time to Nursing Home Placement. *Medicare & Medicaid Research Review*. 2(3).

Santé publique France. *Données hospitalières relatives à l'épidémie de COVID-19*. 30 avril 2020. [<https://www.data.gouv.fr/fr/datasets/donnees-hospitalieres-relatives-a-lepidemie-de-covid-19/>]. Consulté le 30 avril 2020.

Santé publique France. (2020). *Point épidémiologique régional Spécial COVID-19, Hauts-de-France*. 28 mai 2020. [<https://www.santepubliquefrance.fr/regions/hauts-de-france/documents/bulletin-regional/2020/covid-19-point-epidemiologique-en-hauts-de-france-du-28-mai-2020>]. Consulté le 5 juillet 2020.

Silbert, N. (2020). Les auxiliaires de vie, parents pauvres de la lutte contre l'épidémie. *Les échos*. p. 4

Steg, P G. (2020). Les leçons d'une épidémie. *Les échos*. p. 11

Terra. S. (2018). Projections de population dans les Hauts-de-France. *INSEE Flash, Hauts-de-France*. N°50.

Wooldridge, J. (2009). Pooling Crosss Sections across Time : Simple Panel Data Methods. *Introductory Econometrics : A Modern Approach* (4^e édition, p.444-456). Etats-Unis. SOUTH-WESTERN CENGAGE Learning [https://files.transtutors.com/cdn/uploadassignments/386086_5_introductory-econometrics--a-modern-approach-4th---wooldridge.pdf]. Consulté le 3 août 2020.

IX. Annexes

Annexe 1 : enquête auprès des structures

Le mail :

Chers adhérents,

Comme évoqué lors de la visioconférence du 4 juin dernier, dans le but de valoriser l'activité de notre secteur, il nous a semblé intéressant d'évaluer l'impact social des structures de l'aide à domicile pendant la crise sanitaire du Covid-19.

Vous trouverez en pièce jointe un document Excel qui vous permettra de renseigner le nombre de décès et d'hospitalisations sur la période de janvier 2019 à mai 2020.

Vous pouvez également nous transmettre directement les extractions de votre système d'informations.

Nous avons bien conscience que la période reste difficile et que vous êtes souvent sollicités. Néanmoins, **nous vous demandons de répondre avant le 30 juin à un maximum d'éléments**. Il s'agit pour nous d'observer l'efficacité du secteur de l'aide à domicile.

Bonne réception,

Bien à vous

Enquête sur les décès :

[illegible]

Enquête sur les hospitalisations :

[illegible]

Autres annexes

Variables explicatives	Définition
panel	Caractère correspondant à l'année et au mois de décès « Y-M » compris entre janvier 2019 et mai 2020
SEXE	Caractère correspondant au sexe de l'individu. Il prend la valeur M pour les hommes et la valeur F pour les femmes
age	Variable numérique qui correspond à l'âge de la personne décédée
rev_median	Variable numérique qui correspond au revenu médian disponible par unité de consommation selon les communes du Nord
structures	Variable caractère qui correspond à la structure d'aide à domicile qui a répondu à l'enquête
genre	Variable dichotomique qui correspond au sexe de l'individu. Cette variable prend la valeur 1 pour les femmes et 0 pour les hommes
Confinement	Variable dichotomique définit comme la période du confinement. Cette période a été définie comme le mois de mars, avril et mai 2020. Elle prend la valeur 1 sur cette période et 0 sur les autres
Commune	Caractère qui correspond à la commune de résidence de l'individu
population	Variable numérique qui correspond à la population présente dans chaque commune du nord
categorie_commune	Facteur avec trois niveaux : moyenne, riche et pauvres Moyenne : commune avec un revenu médian compris entre 19500€ et 25700€ Riche : commune ayant un revenu médian supérieur à 25700€ Pauvres : commune ayant un revenu médian inférieur à 19500€
classe_ville_pop	Facteur avec quatre niveaux : petite ville, village, ville et grande ville. Village : commune ayant moins de 5000 habitants Petite ville : commune ayant une population qui se situe entre 5000 et 20000 habitants

	Ville : commune ayant une population comprise entre 20000 et 100 000 habitants Grande ville : commune de plus de 100000 habitants
saison	Facteur avec 4 niveaux : hiver, printemps, été, automne Hiver : mois de décembre, janvier et février Printemps : mois de mars, avril et mai Été : mois de juin, juillet et août Automne : mois de septembre, octobre et novembre
nb.deces	Variable numérique qui correspond au nombre de décès par mois
log_deces	Logarithme népérien de la variable nb.deces
log_age	Logarithme népérien de la variable age
communes_riches	Variable indicatrice qui prend la valeur 1 quand categorie_commune = riche et 0 sinon
communes_moyennes	Variable indicatrice qui prend la valeur 1 quand categorie_commune = moyenne et 0 sinon
communes_pauvres	Variable indicatrice qui prend la valeur 1 quand categorie_commune = pauvre et 0 sinon
villages	Variable indicatrice qui prend la valeur 1 quand classe_ville_pop = village et 0 sinon
petites_villes	Variable indicatrice qui prend la valeur 1 quand classe_ville_pop = petite ville et 0 sinon
villes	Variable indicatrice qui prend la valeur 1 quand classe_ville_pop = ville et 0 sinon
grandes_villes	Variable indicatrice qui prend la valeur 1 quand classe_ville_pop = grande ville et 0 sinon
automne	Variable dichotomique qui prend la valeur 1 quand saison = automne et 0 sinon
ete	Variable dichotomique qui prend la valeur 1 quand saison = été et 0 sinon
printemps	Variable dichotomique qui prend la valeur 1 quand saison = printemps et 0 sinon
hiver	Variable dichotomique qui prend la valeur 1 quand saison = hiver et 0 sinon
GIR2	Facteur qui correspond au niveau de dépendance des individus. Il est compris entre 1 et 6.

Tableau 9 : Dictionnaire des variables de la base décès

Variables explicatives	Définition
panel	Caractère correspondant à l'année et au mois d'entrée à l'hôpital « Y-M » compris entre janvier 2019 et mai 2020
Sexe	Caractère correspondant au sexe de l'individu. Il prend la valeur H pour les hommes et la valeur F pour les femmes
age	Variable numérique qui correspond à l'âge de la personne hospitalisée
GIR2	Facteur qui correspond au niveau de dépendance des individus. Il est compris entre 1 et 6.
rev_median	Variable numérique qui correspond au revenu médian disponible par unité de consommation selon les communes du Nord
structures	Variable caractère qui correspond à la structure d'aide à domicile qui a répondu à l'enquête
genre	Variable dichotomique qui correspond au sexe de l'individu. Cette variable prend la valeur 1 pour les femmes et 0 pour les hommes
confinement	Variable dichotomique définit comme la période du confinement. Cette période a été définie comme le mois de mars, avril et mai

	2020. Elle prend la valeur 1 sur cette période et 0 sur les autres
Communes	Caractère qui correspond à la commune de résidence de l'individu
tps_hospi	Variable numérique qui correspond à la durée du séjour à l'hôpital en jour
population	Variable numérique qui correspond à la population présente dans chaque commune du nord
categorie_commune	Facteur avec trois niveaux : moyenne, riche et pauvres Moyenne : commune avec un revenu médian compris entre 19500€ et 25700€ Riche : commune ayant un revenu médian supérieur à 25700€ Pauvres : commune ayant un revenu médian inférieur à 19500€
classe_ville_pop	Facteur avec quatre niveaux : petite ville, village, ville et grande ville : Village : commune ayant moins de 5000 habitants Petite ville : commune ayant une population qui se situe entre 5000 et 20000 habitants Ville : commune ayant une population comprise entre 20000 et 100 000 habitants Grande ville : commune de plus de 100000 habitants
saison	Facteur avec 4 niveaux : hiver, printemps, été, automne Hiver : mois de décembre, janvier et février Printemps : mois de mars, avril et mai Été : mois de juin, juillet et août Automne : mois de septembre, octobre et novembre
nb_hospitalisation	Variable numérique qui correspond au nombre d'entrée à l'hôpital par mois
log_hospi	Logarithme népérien de la variable nb_hospitalisation
log_age	Logarithme népérien de la variable age
log_tps	Logarithme népérien de la variable tps_hospi
communes_riche	Variable indicatrice qui prend la valeur 1 quand categorie_commune = riche et 0 sinon
communes_moyennes	Variable indicatrice qui prend la valeur 1 quand categorie_commune = moyenne et 0 sinon
communes_pauvres	Variable indicatrice qui prend la valeur 1 quand categorie_commune = pauvre et 0 sinon
villages	Variable indicatrice qui prend la valeur 1 quand classe_ville_pop = village et 0 sinon
petites_villes	Variable indicatrice qui prend la valeur 1 quand classe_ville_pop = petite ville et 0 sinon
villes	Variable indicatrice qui prend la valeur 1 quand classe_ville_pop = ville et 0 sinon
grandes_villes	Variable indicatrice qui prend la valeur 1 quand classe_ville_pop = grande ville et 0 sinon
automne	Variable dichotomique qui prend la valeur 1 quand saison = automne et 0 sinon
ete	Variable dichotomique qui prend la valeur 1 quand saison = été et 0 sinon
printemps	Variable dichotomique qui prend la valeur 1 quand saison = printemps et 0 sinon
hiver	Variable dichotomique qui prend la valeur 1 quand saison = hiver et 0 sinon

Tableau 10 : Dictionnaire des variables de la base hospitalisations

L'analyse des échantillons

Evolutions des statistiques descriptives				
Variables	Avant le confinement (N=494)	Confinement (N=126)	Total (N=620)	p-value
age				0.641
Mean (Ecart type)	80.528 (12.884)	81.127 (12.651)	80.650 (12.829)	
genre				0.993
Mean (Ecart type)	0.579 (0.494)	0.579 (0.496)	0.579 (0.494)	
nb.deces				< 0.001
Mean (Ecart type)	37.320 (8.864)	46.063 (12.465)	39.097 (10.313)	
communes_riche				0.098
Mean (Ecart type)	0.026 (0.160)	0.056 (0.230)	0.032 (0.177)	
communes_moyennes				0.985
Mean (Ecart type)	0.451 (0.498)	0.452 (0.500)	0.452 (0.498)	
communes_pauvres				0.546
Mean (Ecart type)	0.522 (0.500)	0.492 (0.502)	0.516 (0.500)	
villages				0.997
Mean (Ecart type)	0.206 (0.405)	0.206 (0.406)	0.206 (0.405)	
petites_villes				0.790
Mean (Ecart type)	0.330 (0.471)	0.317 (0.467)	0.327 (0.470)	
villes				0.553
Mean (Ecart type)	0.267 (0.443)	0.294 (0.457)	0.273 (0.446)	
grandes_villes				0.727
Mean (Ecart type)	0.196 (0.398)	0.183 (0.388)	0.194 (0.395)	

Tableau 11 : Comparaison de l'échantillon des décès avant et pendant le confinement

Matrice des corrélations															
	age	genre	confinement	communes_riche	communes_moyennes	communes_pauvres	villages	petites_villes	villes	grandes_villes	automne	ete	printemps	hiver	nb.deces
age	1	0.198	0.019	0.022	0.060	-0.067	0.074	0.047	-0.045	-0.081	-0.037	0.062	0.002	-0.021	-0.029
genre	0.198	1	-0.002	0.007	0.063	-0.065	0.021	0.023	-0.001	-0.047	-0.024	0.049	-0.027	0.008	0.046
confinement	0.019	-0.002	1	0.066	0.003	-0.026	-0.001	-0.008	0.023	-0.015	-0.195	-0.208	0.660	-0.376	0.341
communes_riche	0.022	0.007	0.066	1	-0.165	-0.190	-0.026	-0.029	0.134	-0.090	0.011	-0.049	0.030	-0.002	-0.006
communes_moyennes	0.060	0.063	0.003	-0.165	1	-0.937	0.244	0.330	-0.175	-0.444	-0.039	0.037	0.034	-0.034	-0.015
communes_pauvres	-0.067	-0.065	-0.026	-0.190	-0.937	1	-0.234	-0.319	0.127	0.474	0.034	-0.019	-0.044	0.034	0.017
villages	0.074	0.021	-0.001	-0.026	0.244	-0.234	1	-0.355	-0.314	-0.251	-0.019	-0.005	0.021	-0.005	-0.036
petites_villes	0.047	0.023	-0.008	-0.029	0.330	-0.319	-0.355	1	-0.426	-0.341	-0.0002	0.010	-0.032	0.025	-0.006
villes	-0.045	-0.001	0.023	0.134	-0.175	0.127	-0.314	-0.426	1	-0.301	-0.042	0.007	0.040	-0.016	0.019
grandes_villes	-0.081	-0.047	-0.015	-0.090	-0.444	0.474	-0.251	-0.341	-0.301	1	0.067	-0.015	-0.029	-0.006	0.022
automne	-0.037	-0.024	-0.195	0.011	-0.039	0.034	-0.019	-0.0002	-0.042	0.067	1	-0.158	-0.296	-0.287	-0.441
ete	0.062	0.049	-0.208	-0.049	0.037	-0.019	-0.005	0.010	0.007	-0.015	-0.158	1	-0.315	-0.305	-0.348
printemps	0.002	-0.027	0.660	0.030	0.034	-0.044	0.021	-0.032	0.040	-0.029	-0.296	-0.315	1	-0.570	0.148
hiver	-0.021	0.008	-0.376	-0.002	-0.034	0.034	-0.005	0.025	-0.016	-0.006	-0.287	-0.305	-0.570	1	0.415
nb.deces	-0.029	0.046	0.341	-0.006	-0.015	0.017	-0.036	-0.006	0.019	0.022	-0.441	-0.348	0.148	0.415	1

Tableau 12 : Matrice des corrélations sur les décès

Evolutions des statistiques descriptives				
Variables	Avant le confinement (N=2421)	Confinement (N=387)	Total (N=2808)	p- value
age				0.867
Mean (Ecart-type)	79.857 (12.700)	79.974 (12.612)	79.873 (12.686)	
genre				0.312
Mean (Ecart-type)	0.672 (0.470)	0.646 (0.479)	0.668 (0.471)	
nb_hospitalisation				< 0.001
Mean (Ecart-type)	175.826 (24.330)	131.977 (19.394)	169.783 (28.118)	
tps_hospi				0.371
Mean (Ecart-type)	127.646 (1481.008)	200.517 (1520.897)	137.689(1486.506)	
communes_riches				0.103
Mean (Ecart-type)	0.027 (0.162)	0.013 (0.113)	0.025 (0.156)	
communes_moyennes				0.869
Mean (Ecart-type)	0.490 (0.500)	0.486 (0.501)	0.490 (0.500)	
communes_pauvres				0.501
Mean (Ecart-type)	0.483 (0.500)	0.501 (0.501)	0.485 (0.500)	
villages				0.304
Mean (Ecart-type)	0.206 (0.405)	0.183 (0.388)	0.203 (0.402)	
petites_villes				0.821
Mean (Ecart-type)	0.351 (0.477)	0.357 (0.480)	0.351 (0.478)	
villes				0.251
Mean (Ecart-type)	0.234 (0.424)	0.261 (0.440)	0.238 (0.426)	
grandes_villes				0.651
Mean (Ecart-type)	0.209 (0.407)	0.199 (0.400)	0.208 (0.406)	

Tableau 13 : Comparaison de l'échantillon d'hospitalisations avant et pendant le confinement

	Matrice des corrélations															
	tps_hospi	age	genre	confinement	communes_riche	communes_moyennes	communes_pauvres	villages	petites_villes	villes	grandes_villes	automne	ete	printemps	hiver	nb_hospitalisation
tps_hospi	1	0.004	0.018	0.017	-0.005	-0.036	0.038	-0.037	-0.025	0.042	0.021	-0.026	-0.008	-0.013	0.040	0.026
age	0.004	1	0.230	0.003	-0.014	0.077	-0.073	0.073	0.066	-0.063	-0.082	-0.008	-0.033	-0.009	0.042	0.044
genre	0.018	0.230	1	-0.015	0.030	0.040	-0.049	-0.017	0.048	0.021	-0.061	0.019	-0.029	-0.008	0.017	0.010
confinement	0.017	0.003	-0.015	1	-0.031	-0.002	0.011	-0.019	0.005	0.019	-0.008	-0.179	-0.185	0.574	-0.280	-0.537
communes_riche	-0.005	-0.014	0.030	-0.031	1	-0.156	-0.157	0.036	-0.076	0.131	-0.083	-0.004	-0.008	0.015	-0.005	0.014
communes_moyennes	-0.036	0.077	0.040	-0.002	-0.156	1	-0.951	0.184	0.431	-0.177	-0.501	-0.021	-0.034	0.024	0.019	0.015
communes_pauvres	0.038	-0.073	-0.049	0.011	-0.157	-0.951	1	-0.195	-0.407	0.136	0.527	0.022	0.036	-0.029	-0.017	-0.019
villages	-0.037	0.073	-0.017	-0.019	0.036	0.184	-0.195	1	-0.367	-0.278	-0.255	0.030	-0.018	-0.005	-0.004	-0.007
petites_villes	-0.025	0.066	0.048	0.005	-0.076	0.431	-0.407	-0.367	1	-0.415	-0.381	-0.029	-0.016	0.029	0.007	0.005
villes	0.042	-0.063	0.021	0.019	0.131	-0.177	0.136	-0.278	-0.415	1	-0.288	-0.004	0.010	0.002	-0.008	-0.005
grandes_villes	0.021	-0.082	-0.061	-0.008	-0.083	-0.501	0.527	-0.255	-0.381	-0.288	1	0.009	0.026	-0.032	0.003	0.006
automne	-0.026	-0.008	0.019	-0.179	-0.004	-0.021	0.022	0.030	-0.029	-0.004	0.009	1	-0.207	-0.311	-0.314	-0.210
ete	-0.008	-0.033	-0.029	-0.185	-0.008	-0.034	0.036	-0.018	-0.016	0.010	0.026	-0.207	1	-0.322	-0.325	-0.047
printemps	-0.013	-0.009	-0.008	0.574	0.015	0.024	-0.029	-0.005	0.029	0.002	-0.032	-0.311	-0.322	1	-0.489	-0.295
hiver	0.040	0.042	0.017	-0.280	-0.005	0.019	-0.017	-0.004	0.007	-0.008	0.003	-0.314	-0.325	-0.489	1	0.498
nb_hospitalisation	0.026	0.044	0.010	-0.537	0.014	0.015	-0.019	-0.007	0.005	-0.005	0.006	-0.210	-0.047	-0.295	0.498	1

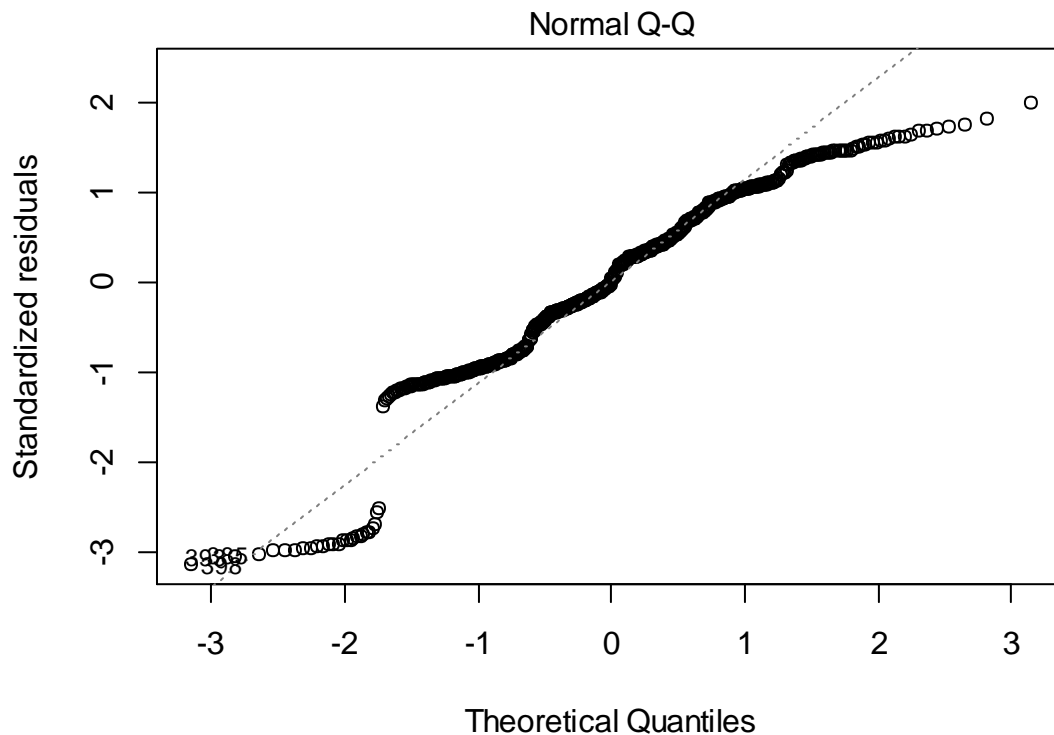
Tableau 14 : Matrice des corrélations entre les variables des hospitalisations

tableau d'analyse des variances

	Df	Somme des carrés	Moyenne des carrés	F value	Pr(> F)
log_age	1	0.056	0.056	1.701	0.193
genre	1	0.117	0.117	3.527	0.061
communes_riches	1	0.007	0.007	0.217	0.641
communes_pauvres	1	0.010	0.010	0.290	0.590
villages	1	0.020	0.020	0.601	0.439
petites_villes	1	0.012	0.012	0.367	0.545
villes	1	0.000	0.000	0.0001	0.994
hiver	1	8.958	8.958	269.946	0
printemps	1	10.143	10.143	305.638	0
ete	1	0.465	0.465	14.021	0.0002
Confinement	1	3.311	3.311	99.779	0
Residuals	606	20.111	0.033		

Tableau 15 : Tableau d'analyse des variances sur les décès

L'analyse du modèle économétrique



$\text{lm}(\log_deces \sim \log_age + \text{genre} + \text{communes_riche} + \text{communes_pauvres} + \text{villa}$

Figure 6 : QQplot des résidus

Test de normalité de Wilk-Shapiro		
	MCO	MCG
Equation des décès	0	0

Tableau 16 : p-value des tests de normalité

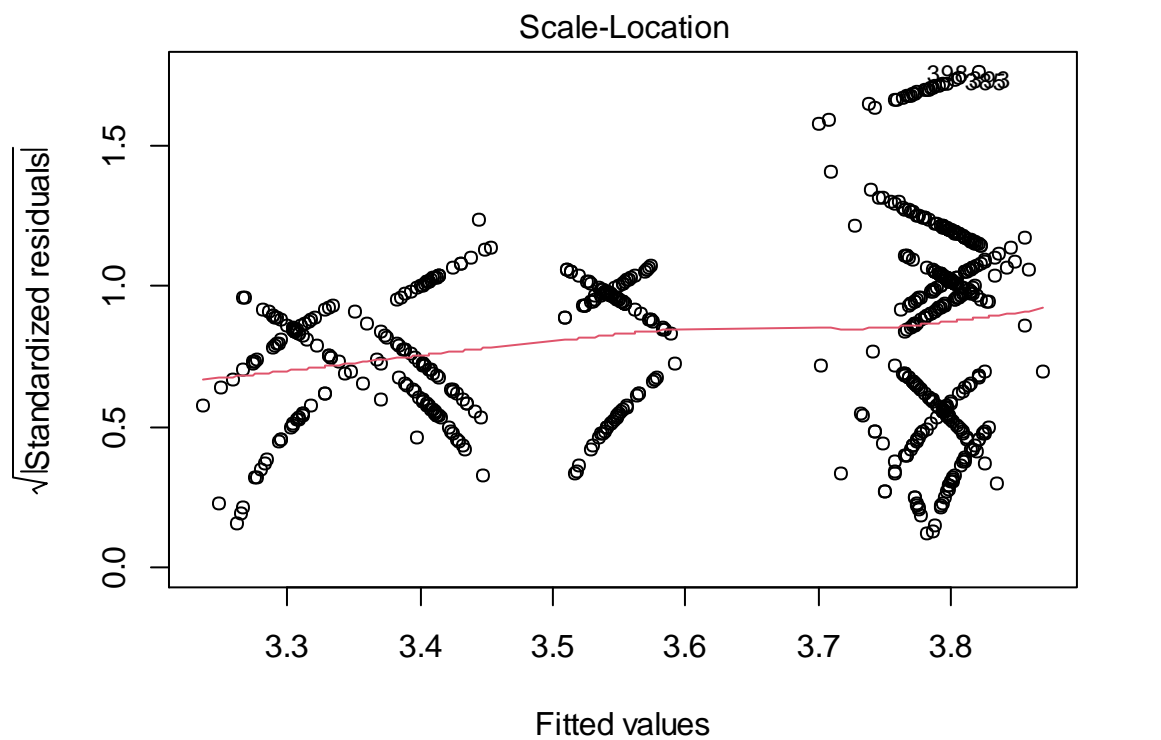


Figure 7 : Détection graphique de l'hétéroscédasticité

Test de Breusch-Pagan	
Equation des décès	0

Tableau 17 : p-value du test de Breusch-Pagan de détection de l'hétéroscédasticité

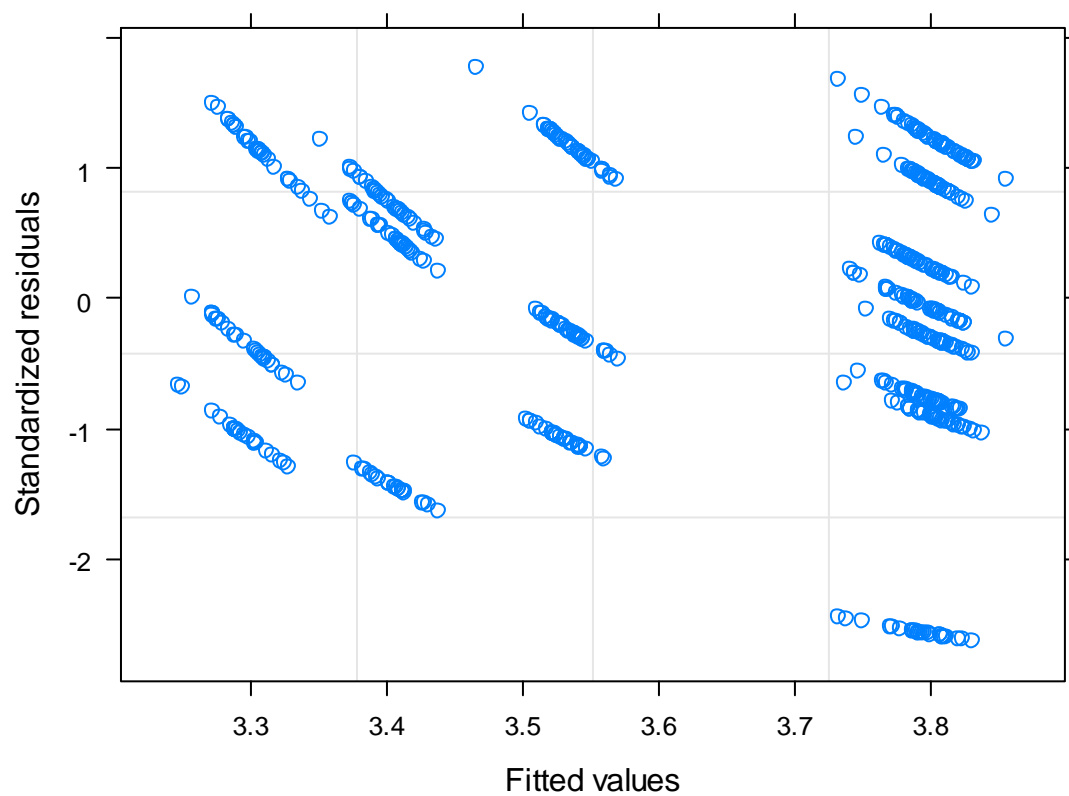


Figure 8 : Graphique après correction de l'hétéroscédasticité par la méthode des MCG

Test de Durbin-Watson	
Equation des décès	1.4275

Tableau 18 : Statistique de test du test de Durbin-Watson pour détecter l'autocorrélation

autocorrélogramme des résidus

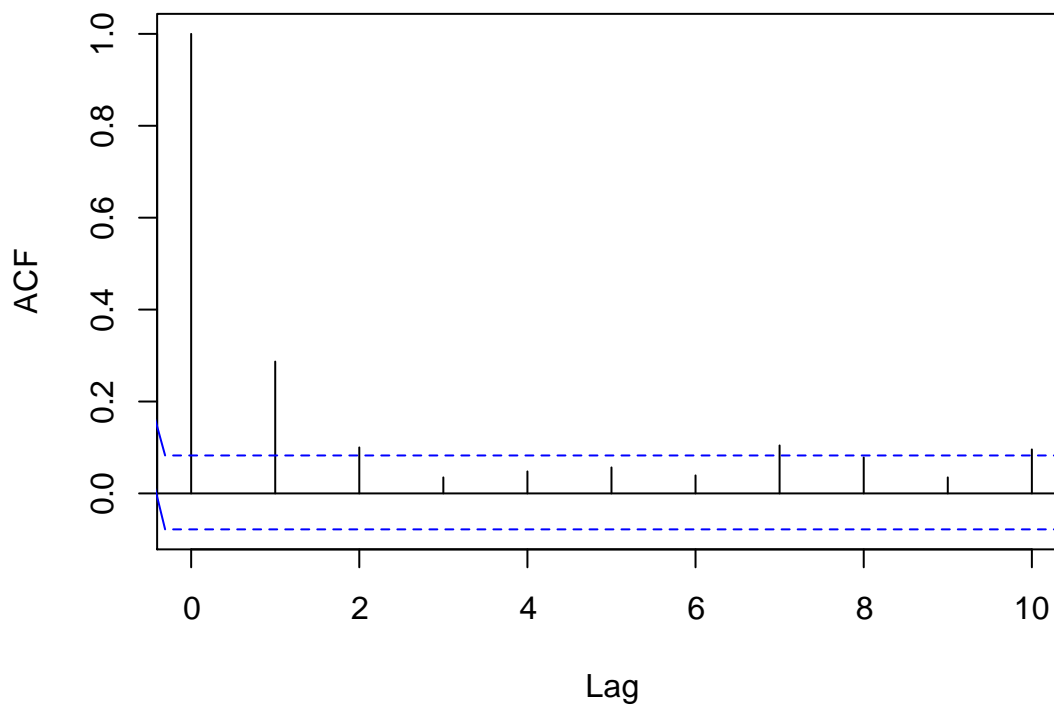


Figure 9 : Autocorrélogramme des résidus du modèle MCO sur les décès

test de multi-colinéarité	
Variables explicatives	VIF
log_age	1.049
genre	1.035
communes_riches	1.08
communes_pauvres	1.518
villages	2.145
petites_villes	2.473
viles	1.932
hiver	2.426
printemps	3.229
ete	1.831
Confinement	1.78

Tableau 19 : Test de multi-colinéarité entre les variables

résultat d'estimation sur les décès de la MAD

	Variable dépendante :		
	Log_deces		
	MCO	MCG	MCG corrigé
log_age	-0.0187 (0.0499)	-0.0187 (0.0499)	-0.0030 (0.0440)
genre	0.0262 (0.0192)	0.0262 (0.0192)	0.0007 (0.0162)
communes_riches	-0.1220** (0.0550)	-0.1220** (0.0550)	-0.1334*** (0.0416)
communes_pauvres	-0.0430 (0.0306)	-0.0430 (0.0306)	-0.0502* (0.0266)
villages	-0.0111 (0.0512)	-0.0111 (0.0512)	-0.0011 (0.0454)
petites_villes	-0.0592 (0.0381)	-0.0592 (0.0381)	-0.0642** (0.0325)
viles	-0.0277 (0.0270)	-0.0277 (0.0270)	-0.0329 (0.0228)
ete	-0.0617* (0.0316)	-0.0617* (0.0316)	-0.0561*** (0.0216)
printemps	0.1076*** (0.0303)	0.1076*** (0.0303)	0.1093*** (0.0237)
hiver	0.3054*** (0.0275)	0.3054*** (0.0275)	0.3078*** (0.0246)
confinement	0.2591*** (0.0341)	0.2591*** (0.0341)	0.2761*** (0.0415)
Constant	3.1845*** (0.2193)	3.1845*** (0.2193)	3.1339*** (0.1926)
Observations	692	692	692
R²	0.3061		
Adjusted R ²	0.2949		
Log Likelihood		-32.7580	25.4815
Akaike Inf. Crit.		91.5160	-20.9631
Bayesian Inf. Crit.		150.3032	46.8684
Residual Std. Error	0.2442 (df = 680)		
F Statistic	27.2705*** (df = 11; 680)		

Note:

* p<0.1; ** p<0.05; *** p<0.01

Tableau 20 : Résultats des estimations des décès de la MAD depuis 2018

Comparaison des résultats d'estimation sur les décès			
	<i>Variable dépendante:</i>		
	Log_deces		
	MCG MAD	MCG Nord	MCG Hauts-de-France
log_age	0.0666 (0.0719)	0.0070*** (0.0024)	0.0071*** (0.0016)
genre	-0.0030 (0.0185)	0.0002 (0.0006)	0.00004 (0.0004)
communes_riche	-0.1401*** (0.0488)	0.0003 (0.0024)	-0.0019 (0.0017)
communes_pauvres	-0.0535* (0.0299)	0.0001 (0.0009)	0.0002 (0.0006)
villages	-0.0011 (0.0499)	0.0012 (0.0012)	0.0006 (0.0008)
petites_villes	-0.0734** (0.0365)	0.0003 (0.0010)	0.0003 (0.0007)
villes	-0.0188 (0.0261)	0.0002 (0.0009)	0.00001 (0.0006)
ete	0.0174 (0.0259)	-0.0470*** (0.0007)	-0.0599*** (0.0005)
printemps	0.1175*** (0.0266)	0.0927*** (0.0011)	0.0890*** (0.0007)
hiver	0.2991*** (0.0262)	0.1639*** (0.0013)	0.1479*** (0.0007)
confinement	0.2916*** (0.0412)	0.0591*** (0.0021)	0.1071*** (0.0015)
Constant	2.7221*** (0.3149)	7.3346*** (0.0106)	8.2178*** (0.0069)
Observations	634	49,126	118,497
Log Likelihood	5.6872	50,137.8500	125,924.9000
Akaike Inf. Crit.	18.6255	-100,247.7000	-251,821.9000
Bayesian Inf. Crit.	85.1196	-100,124.5000	-251,686.3000

Note: * p<0.1; ** p<0.05; *** p<0.01

Tableau 21 : Comparaison des décès sur trois populations différentes

Variables	MAD	Nord	p-value
age	82.67	81.23	0.0003
genre	59.78%	52.78%	0.0004
confinement	11.67%	11.03%	0.6166
communes_riche	3%	1.85%	0.0919
communes_moyennes	46.85%	17.88%	0
communes_pauvres	50.16%	80.27%	0
villages	5.99%	13.83%	0
petites_villes	28.71%	26.14%	0.1569
villes	33.44%	42.03%	0
grandes_villes	31.86%	18%	0
automne	17.51%	19.28%	0.2441
ete	18.45%	18.42%	0.9803
printemps	31.86%	32.02%	0.9305
hiver	32.18%	30.28%	0.3102
nb_deces	23.68	1722.61	0

Tableau 22 : Comparaison de la population des personnes décédées dans le Nord et dans la MAD sur la population des plus de soixante ans

test de multi-colinéarité

Variables explicatives **VIF**

log_age	1.074
genre	1.057
communes_riche	1.067
communes_pauvres	1.658
villages	2.133
petites_villes	2.700
villes	1.819
ete	1.699
hiver	2.000
printemps	2.49
log_tps	1.039
confinement	1.498

Tableau 23 : Test de multi-colinéarité entre les variables explicatives des hospitalisations

tableau d'analyse des variances

	Df	Somme des carrés	Moyenne des carrés	F value	Pr(> F)
log_age	1	0.137	0.137	10.277	0.001
genre	1	0.000	0.000	0.0001	0.994
communes_riches	1	0.023	0.023	1.701	0.192
communes_pauvres	1	0.013	0.013	0.944	0.331
villages	1	0.006	0.006	0.464	0.496
petites_villes	1	0.009	0.009	0.678	0.411
villes	1	0.024	0.024	1.790	0.181
ete	1	0.035	0.035	2.601	0.107
hiver	1	18.778	18.778	1,409.427	0
printemps	1	0.024	0.024	1.792	0.181
log_tps	1	0.0002	0.0002	0.017	0.898
confinement	1	19.763	19.763	1,483.405	0
Residuals	2,767	36.865	0.013		

Tableau 24 : Tableau d'analyse des variances sur les hospitalisations

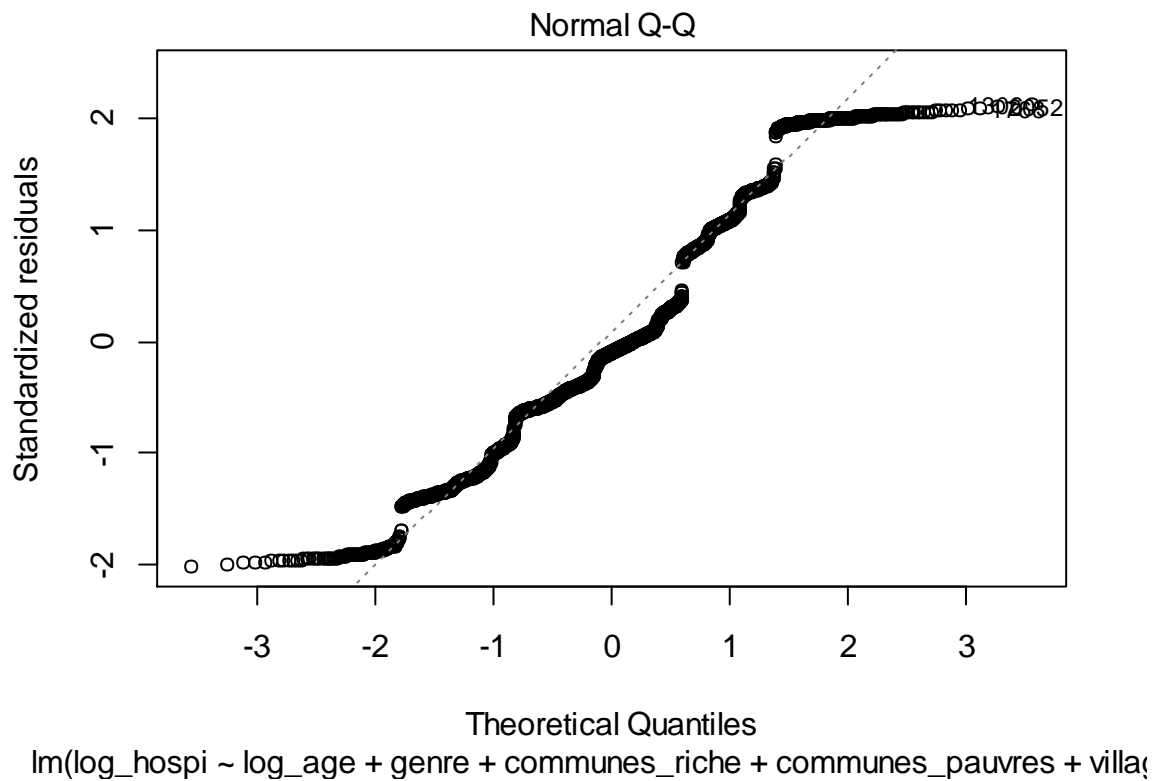


Figure 10 : QQplot des résidus

Test de normalité de Wilk-Shapiro		
	MCO	MCG
Equation des hospitalisations	0	0

Tableau 25 : p-value des tests de normalité

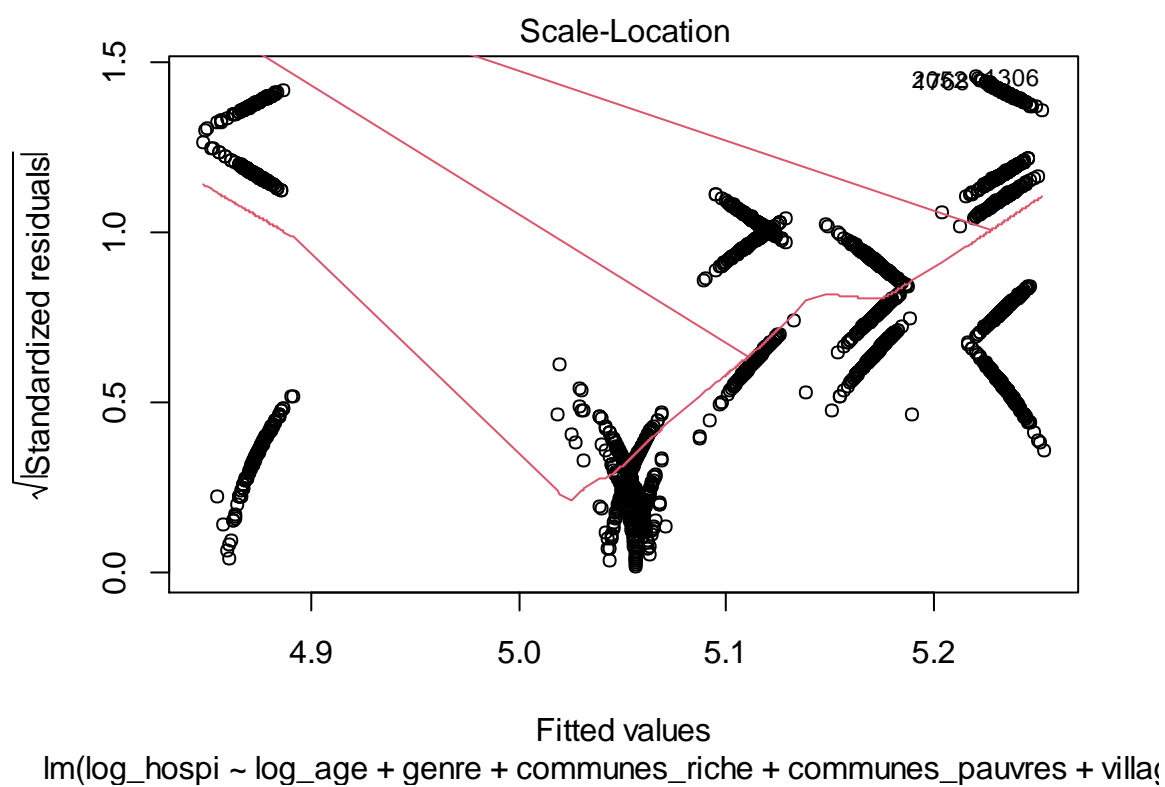


Figure 11 : Détection graphique de l'hétéroscédasticité

Test de Breusch-Pagan	
Equation des décès	0

Tableau 26 : p-value du test de Breusch-Pagan de détection de l'hétéroscédasticité

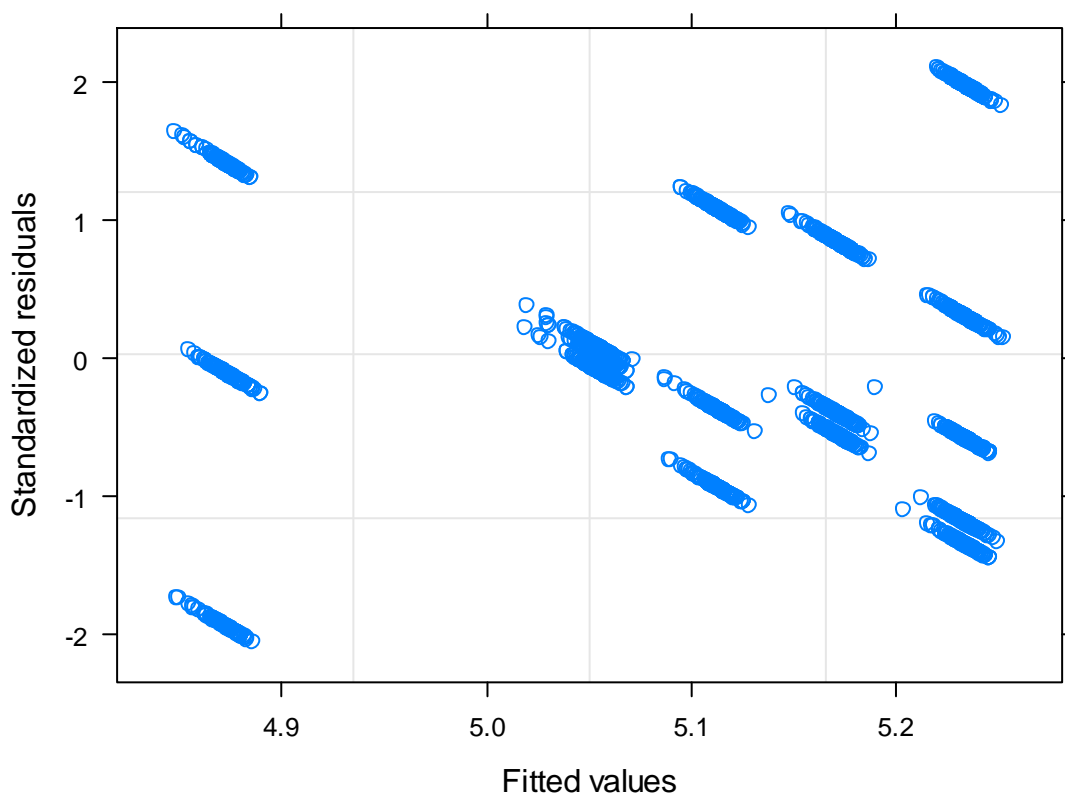


Figure 12 : Graphique après correction de l'hétéroscédasticité par la méthode des MCG

Test de Durbin-Watson	
Equation des décès	1.91

Tableau 27: Statistique du test de Durbin-Watson de détection de l'autocorrélation

résultat d'estimation sur les hospitalisations

	Variable dépendante:			
	Nb_hospitalisation			
	MCO	MCG	MCG corrigé	correction de white
age	0.0758** (0.0318)	0.0758** (0.0318)	0.0647** (0.0294)	0.0758** (0.0302)
genre	-0.5017 (0.8529)	-0.5017 (0.8529)	-0.6823 (0.7970)	-0.5017 (0.8475)
communes_riche	-0.9304 (2.5647)	-0.9304 (2.5647)	-0.7589 (2.4694)	-0.9304 (2.4157)
communes_pauvres	-0.5189 (1.008)	-0.5189 (1.008)	-0.3005 (0.9310)	-0.5189 (0.9615)
villages	-1.5085 (1.4115)	-1.5085 (1.4115)	-1.4617 (1.3240)	-1.5085 (1.3437)
petites_villes	-1.0336	-1.0336	-1.2432	-1.0336

	(1.3315)	(1.3315)	(1.2386)	(1.3140)
villes	-0.1571	-0.1571	-0.5905	-0.1571
	(1.2306)	(1.2306)	(1.1463)	(1.2559)
ete	10.2622***	10.2622***	10.2372***	10.2622***
	(1.3297)	(1.3297)	(1.2080)	(0.7700)
hiver	32.8215***	32.8215***	32.8388***	32.8215***
	(1.1709)	(1.1709)	(1.1298)	(0.9733)
printemps	20.0125***	20.0125***	20.0269***	20.0125***
	(1.3095)	(1.3095)	(1.2385)	(0.5942)
tps_hospi	0.0003*	0.0003*	0.0002	0.0003
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
confinement	-44.6820***	-44.6820***	-44.6653***	-44.6820***
	(1.3831)	(1.3831)	(1.1709)	(1.1619)
Constant	151.9657***	151.9657***	153.0364***	151.9657***
	(2.8814)	(2.8814)	(2.6514)	(2.6213)
Observations	2,780	2,780	2,780	
R²	0.4659			
Adjusted R ²	0.4636			
Log Likelihood		-12,336.7	-12,229.53	
Akaike Inf. Crit.		24,701.41	24,489.07	
F Statistic	201.16*** (df = 12; 2735)			

Note: () : Ecart-type ; * p<0.1; ** p<0.05; *** p<0.01

Tableau 28 : Estimation du nombre d'hospitalisations en niveau-niveau

L'ajout du GIR dans l'échantillon

Le niveau de dépendance d'un individu est approximé par le GIR. Il existe 6 niveaux de dépendance. Cette variable n'est pas analysée dans l'étude précédente car pour beaucoup d'individus cette donnée est manquante. Néanmoins, voici la répartition des individus selon le GIR pour le décès et pour les hospitalisations :

Niveau de GIR	Décès		Hospitalisations	
	Nombre	répartition	Nombre	Répartition
GIR 1 (très dépendant)	18	2.91%	19	0.68%
GIR 2 (très dépendant)	95	15.32%	352	12.54%
GIR 3 (dépendant)	101	16.29%	643	22.9%
GIR 4 (dépendant)	174	28.06%	964	34.33%
GIR 5 (autonome)	43	6.94%	241	8.58%
GIR 6 (autonome)	85	13.71%	282	10.04%
NA	104	16.77%	307	10.93%
total	620	100%	2808	100%

Tableau 29 : répartition des individus selon le GIR dans les deux bases de données

Dans les deux échantillons les **personnes dépendantes et très dépendantes représentent la majorité des individus soit 62.58% pour les décès et 70.45% pour les hospitalisations**. L'introduction du GIR supprime 2 structures de la base de décès et 1 structure de la base des hospitalisations.

Comparaison du modèle sur les décès avec ou sans le GIR

résultat d'estimation sur les décès				
	<i>Dependent variable:</i>			
	Log_deces		MCG	MCG gir
	MCO	MCO gir		
log_age	-0.0540 (0.0408)	-0.0696 (0.0454)	-0.0374 (0.0348)	-0.0522 (0.0394)
genre	0.0278* (0.0151)	0.0370** (0.0170)	0.0161 (0.0123)	0.0226 (0.0138)
GIR22		-0.0775 (0.0482)		-0.0446 (0.0425)
GIR23		-0.0453 (0.0480)		-0.0203 (0.0420)
GIR24		-0.0559 (0.0466)		-0.0413 (0.0411)
GIR25		-0.0609 (0.0527)		-0.0616 (0.0452)
GIR26		-0.0706 (0.0488)		-0.0553 (0.0429)
communes_riche	-0.0672 (0.0430)	-0.0738 (0.0471)	-0.0558 (0.0357)	-0.0602 (0.0382)
communes_pauvres	-0.0110 (0.0181)	-0.0027 (0.0206)	-0.0183 (0.0161)	-0.0202 (0.0177)
villages	-0.0442* (0.0265)	-0.0500 (0.0311)	-0.0376 (0.0254)	-0.0388 (0.0296)
petites_villes	-0.0350 (0.0246)	-0.0333 (0.0270)	-0.0373* (0.0222)	-0.0448* (0.0237)
villes	-0.0206 (0.0229)	-0.0250 (0.0248)	-0.0389** (0.0195)	-0.0443** (0.0208)
hiver	0.4932*** (0.0238)	0.5019*** (0.0269)	0.4985*** (0.0192)	0.4991*** (0.0212)
printemps	0.2458*** (0.0273)	0.2448*** (0.0315)	0.2346*** (0.0199)	0.2344*** (0.0224)

ete	0.1052*** (0.0282)	0.1037*** (0.0317)	0.1042*** (0.0184)	0.0997*** (0.0202)
confinement	0.2424*** (0.0243)	0.2540*** (0.0275)	0.2687*** (0.0258)	0.2707*** (0.0281)
Constant	3.5550*** (0.1782)	3.6674*** (0.2001)	3.4950*** (0.1512)	3.6002*** (0.1708)
Observations	618	516	618	516
R ²	0.5346	0.5434		
Adjusted R ²	0.5261	0.5287		
Log Likelihood			206.9542	144.9598
Akaike Inf. Crit.			-383.9084	-249.9197
F Statistic	63.2805*** (df = 11; 606)	37.1146*** (df = 16; 499)		
<i>Note:</i>		* p<0.1; ** p<0.05; *** p<0.01		

L'introduction du niveau de dépendance dans le modèle pour expliquer le logarithme des décès ne permet pas d'augmenter de manière importante le R². De plus, lorsque l'on compare les AIC des 2 modèles corrigés de la même façon. On obtient un AIC inférieure dans le cas du modèle sans le GIR. Donc l'introduction du GIR ne permet pas de minimiser le AIC ni de maximiser le R². L'introduction du niveau de dépendance mesuré par **GIR n'apporte que peu de puissance statistique supplémentaire**. Il n'est donc pas nécessaire de l'introduire dans le modèle.

Comparaison du modèle sur les hospitalisations avec ou sans le GIR

résultat d'estimation sur les hospitalisations

	<i>Dependent variable:</i>			
	Log_hospi			
	MCO	MCO GIR	MCG	MCG GIR
log_age	0.0285** (0.0120)	0.0314** (0.0125)	0.0283** (0.0120)	0.0313** (0.0125)
genre	-0.0034 (0.0048)	-0.0023 (0.0051)	-0.0035 (0.0048)	-0.0025 (0.0051)
GIR22		0.0108 (0.0274)		0.0110 (0.0274)
GIR23		0.0038 (0.0272)		0.0038 (0.0272)
GIR24		0.0095 (0.0270)		0.0094 (0.0270)
GIR25		0.0221		0.0221

		(0.0277)		(0.0277)
GIR26		0.0227		0.0233
		(0.0276)		(0.0276)
communes_riche	-0.0054	-0.0031	-0.0053	-0.0029
	(0.0144)	(0.0146)	(0.0145)	(0.0146)
communes_pauvres	-0.0021	-0.0002	-0.0021	0.00003
	(0.0056)	(0.0061)	(0.0056)	(0.0061)
villages	-0.0068	-0.0088	-0.0070	-0.0093
	(0.0080)	(0.0087)	(0.0080)	(0.0087)
petites_villes	-0.0058	-0.0036	-0.0060	-0.0038
	(0.0075)	(0.0079)	(0.0075)	(0.0079)
villes	-0.0010	-0.00004	-0.0013	-0.0003
	(0.0069)	(0.0073)	(0.0069)	(0.0073)
ete	0.0584***	0.0573***	0.0584***	0.0573***
	(0.0075)	(0.0078)	(0.0074)	(0.0077)
hiver	0.1790***	0.1786***	0.1790***	0.1786***
	(0.0066)	(0.0069)	(0.0066)	(0.0069)
printemps	0.1173***	0.1168***	0.1173***	0.1167***
	(0.0074)	(0.0078)	(0.0074)	(0.0078)
log_tps	0.0017	0.0011	0.0016	0.0010
	(0.0013)	(0.0014)	(0.0013)	(0.0014)
confinement	-0.2997***	-0.3003***	-0.2997***	-0.3003***
	(0.0078)	(0.0083)	(0.0077)	(0.0082)
Constant	4.9329***	4.9085***	4.9341***	4.9096***
	(0.0521)	(0.0577)	(0.0520)	(0.0575)
Observations	2,780	2,494	2,780	2,494
R ²	0.5129	0.5112		
Adjusted R ²	0.5107	0.5078		
Log Likelihood			2,009.0830	1,785.6150
Akaike Inf. Crit.			-3,988.1650	-3,531.2300
F Statistic	242.7579*** (df = 12; 2767)	152.2924*** (df = 17; 2476)		

Note:

* p<0.1; ** p<0.05; *** p<0.01

L'introduction du GIR pour estimer les hospitalisations n'est pas plus cohérente. En effet, l'introduction du **GIR n'entraîne une augmentation importante du R²**. De même, le AIC est plus petit sans le GIR qu'avec. L'introduction du GIR ne se justifie pas dans le cadre des hospitalisations ni sur les décès puisque l'augmentation du R² est très faible et que le AIC ne baisse pas après l'introduction du GIR dans le modèle estimé.

Code pour la construction de la base de données

#base de données décès MAD

```
clients_nov19<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/Listes clients par mois/CLIENT NOV19.csv", sep = ";", header = T)
head(clients_nov19)
clients_dec19<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/Listes clients par mois/CLIENT DEC19.csv", sep = ";", header = T)
head(clients_dec19)
clients_jan20<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/Listes clients par mois/CLIENT JAN20.csv", sep = ";", header = T)
head(clients_jan20)
clients_fev20<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/Listes clients par mois/CLIENT FEV20.csv", sep = ";", header = T)
head(clients_fev20)
clients_mar20<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/Listes clients par mois/CLIENT MAR20.csv", sep = ";", header = T)
head(clients_mar20)
clients_avr20<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/Listes clients par mois/CLIENT AVR20.csv", sep = ";", header = T)
head(clients_avr20)
clients_mai20<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/fichier client MAD mai 2020.csv", sep = ";", header = T)
clients_2017_2020<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/analyse DC (1) tot.csv", sep = ";", header = T)
library(tidyverse)
clients_2017_2020 %>%
  select(-RESENCE,-dcd)->clients_2017_2020
```

#Construction de la base MAD 2017 - mai 2020 (décès)

```
clients_2017_2020 %>%
  full_join(clients_nov19)->pan1
pan1 %>%
  full_join(clients_dec19)->pan2
```

```

pan2 %>%
  full_join(clients_jan20)->pan3
pan3 %>%
  full_join(clients_fev20)->pan4
pan4 %>%
  full_join(clients_mar20)->pan5
pan5 %>%
  full_join(clients_avr20)->pan6
pan6 %>%
  full_join(clients_mai20)->pan7

```

#Suppression de doublons

```

pan7 %>%
  filter(Date.décès != « »)->pan7
pan7 %>%
  distinct(Code.Aidé,.keep_all = T)->PanMAD

```

#Sélection des variables pertinentes

```

PanMAD %>%
  select(Code.Aidé, Titre, Date.décès, Date.naissance, GIR, Situation.familiale,
  Commune)->PanMAD
PanMAD %>%
  mutate(Date.naissance=as.Date(Date.naissance, format = "%d/%m/%Y")) %>%
  mutate(Date.décès=as.Date(Date.décès, format = "%d/%m/%Y"))->PanMAD

```

#Construction de la variable âge

```

library(lubridate)
PanMAD %>%
  mutate(age = trunc(time_length(interval(Date.naissance, Date.décès), "years")))-
>PanMAD

```

```

PanMAD$Titre<-as.character(PanMAD$Titre)
PanMAD$Titre[PanMAD$Titre %in% c("ME", "MELLE", "MLE", "MLLE","MME")]<-"F"
PanMAD$Titre[PanMAD$Titre %in% "MR"]<-"H"

```

```

PanMAD %>%
  filter(Titre=="F"|Titre=="H")->PanMAD
PanMAD$GIR<-as.character(PanMAD$GIR)
PanMAD$GIR2[PanMAD$GIR %in% ""]<-" "
PanMAD$GIR[PanMAD$GIR %in% c("1", "1A")]<-"1"
PanMAD$GIR[PanMAD$GIR %in% c("2", "2A", "2B", "2C", "2D")]<-"2"
PanMAD$GIR[PanMAD$GIR %in% c("3", "3E", "3F")]<-"3"
PanMAD$GIR[PanMAD$GIR %in% c("4", "4G", "4H")]<-"4"
PanMAD$GIR[PanMAD$GIR %in% c("5A", "5B", "5C", "5H")]<-"5"

```

```

PanMAD$GIR[PanMAD$GIR %in% c("6", "6A", "6B", "6C", "6H")]<-"6"
PanMAD %>%
  mutate(GIR2=as.factor(GIR))->PanMAD
levels(PanMAD$GIR2)
PanMAD %>% select(-GIR)->PanMAD
PanMAD %>%
  filter(Date.décès>=2017-01-01)->MAD2017_2020

```

#Création des variables genre et couple

```

MAD2017_2020 %>%
  mutate(genre=ifelse(Titre=="F",1,0))->MAD2017_2020
MAD2017_2020$Situation.familiale[MAD2017_2020$Situation.familiale %in% ""]<-" "
MAD2017_2020$Situation.familiale[MAD2017_2020$Situation.familiale %in%
c("Marié(e)", "PACS", "Vie maritale")]<-"1"
MAD2017_2020$Situation.familiale[MAD2017_2020$Situation.familiale %in%
c("Célibataire", "Divorcé", "Séparé", "Veuf")]<-"0"
MAD2017_2020 %>%
  mutate(couple = as.factor(Situation.familiale))->MAD2017_2020

```

#Ajout des données de revenu et de population. Les données ont été retravaillées sur excel (Mise en majuscule des communes et suppression des accents, traits d'union et apostrophes)

```

revenu_communes<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/rev_commune.csv", sep=";", header = T)
MAD2017_2020 %>%
  full_join(revenu_communes, by = c("Commune" = "commune")) %>%
  filter(!is.na(Titre))->MAD2017_2020
pop<-read.csv2("Stage-de-M2/base de données/population communes nord.csv")
MAD2017_2020 %>%
  full_join(pop, by = "Commune") %>%
  filter(!is.na(Titre))->MAD2017_2020

```

#Construction de la dimension temporelle

```

MAD2017_2020 %>%
  mutate(panel = format(Date.décès, format = "%Y-%m"))->MAD2017_2020
MAD2017_2020 %>%
  select(-Date.naissance, -Date.décès)->MAD2017_2020
MAD2017_2020 %>%
  mutate(confinement = ifelse(panel=="2020-03"|panel=="2020-04"|panel=="2020-
05", 1, 0))->MAD2017_2020
MAD2017_2020$panel2[MAD2017_2020$panel %in% c("2017-01", "2017-02", "2017-
12", "2018-01", "2018-02", "2018-12", "2019-01", "2019-02", "2019-12", "2020-01",
"2020-02")]<-"hiver"

```

```

MAD2017_2020$panel2[MAD2017_2020$panel %in% c("2017-03", "2017-04",
"2017-05", "2018-03", "2018-04", "2018-05", "2019-03", "2019-04", "2019-05", "2020-
03", "2020-04", "2020-05")]<-"printemps"
MAD2017_2020$panel2[MAD2017_2020$panel %in% c("2017-06", "2017-07",
"2017-08", "2018-06", "2018-07", "2018-08", "2019-06", "2019-07", "2019-08")]<-"été"
MAD2017_2020$panel2[MAD2017_2020$panel %in% c("2017-09", "2017-10",
"2017-11", "2018-09", "2018-10", "2018-11", "2019-09", "2019-10", "2019-11")]<-
"automne"
MAD2017_2020 %>%
  mutate(saison = as.factor(panel2)) %>%
  select(-panel2)->MAD2017_2020

```

#Construction des catégories des communes

```

MAD2017_2020$categorie_commune<-cut(MAD2017_2020$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes", "riches"))
MAD2017_2020$classe_ville_pop<-cut(MAD2017_2020$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))

```

#Pour plus de facilité dans l'analyse descriptive les facteurs vont être transformés en numérique

```

MAD2017_2020 %>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
  mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
  mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
  mutate(automne = ifelse(saison=="automne", 1, 0)) %>%
  mutate(ete = ifelse(saison=="été", 1, 0)) %>%
  mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%
  mutate(hiver = ifelse(saison=="hiver", 1, 0))->MAD2017_2020
MAD2017_2020 %>%
  group_by(panel) %>%
  mutate(nb.deces=n())->MAD2017_2020
MAD2017_2020 %>%
  filter(!is.na(saison))->MAD2017_2020
MAD2017_2020 %>%
  mutate(log_age = log(age)) %>%
  mutate(log_deces = log(nb.deces))->MAD2017_2020

```

#Construction de la base des structures UNA

#Réduction de l'échantillon de la MAD sur les années 2019 - 2020

```
MAD2017_2020 %>%
  filter(panel>="2019-01") %>%
  select(-Code.Aidé, -Situation.familiale, -couple, -nb.deces,-log_age,-log_deces)
%>%
  mutate(structures = "MAD")->UNA2019_2020
UNA2019_2020 %>%
  rename(Sexe=Titre)->UNA2019_2020
```

#Ajout des décès dans les autres structures (On répète à peu près les mêmes étapes que pour la MAD)

#Bailleul

```
read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage évaluation impact aide
à domicile/données/études des données MAD/Arnaud stat/Arnaud stat/décès 2019-
2020.csv", sep = ";", header = T, skip = 1)->deces_bailleul
deces_bailleul %>%
  select(Sexe, Date.de.décès, age, Communes, GIR2)->deces_bailleul
deces_bailleul %>%
  mutate(date.deces = as.Date(Date.de.décès, format = "%d/%m/%Y")) %>%
  mutate(panel = format(date.deces, format = "%Y-%m")) %>%
  filter(!is.na(age))->deces_bailleul
deces_bailleul %>%
  select(-Date.de.décès, -date.deces)->deces_bailleul
deces_bailleul %>%
  mutate(genre=ifelse(Sexe=="F",1,0)) %>%
  mutate(confinement=ifelse(panel=="2020-03"|panel=="2020-04"|panel=="2020-
05",1,0)) %>%
  mutate(GIR2 = as.factor(as.character(GIR2)))->deces_bailleul
deces_bailleul %>%
  full_join(revenu_communes, by = c("Communes" = "commune")) %>%
  filter(!is.na(panel))->deces_bailleul
deces_bailleul %>%
  inner_join(pop, by = c("Communes"="Commune"))->deces_bailleul
deces_bailleul$panel2[deces_bailleul$panel %in% c("2019-01", "2019-02", "2019-
12", "2020-01", "2020-02")]<-"hiver"
deces_bailleul$panel2[deces_bailleul$panel %in% c("2019-03", "2019-04", "2019-
05", "2020-03", "2020-04", "2020-05")]<-"printemps"
deces_bailleul$panel2[deces_bailleul$panel %in% c("2019-06", "2019-07", "2019-
08")]<-"été"
deces_bailleul$panel2[deces_bailleul$panel %in% c("2019-09", "2019-10", "2019-
11")]<-"automne"
deces_bailleul %>%
```

```

mutate(saison = as.factor(panel2)) %>%
select(-panel2)->deces_bailleul
deces_bailleul$categorie_commune<-cut(deces_bailleul$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes","riches"))
deces_bailleul$classe_ville_pop<-cut(deces_bailleul$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))
deces_bailleul %>%
mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
mutate(automne = ifelse(saison=="automne", 1, 0)) %>%
mutate(ete = ifelse(saison=="été", 1, 0)) %>%
mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%
mutate(hiver = ifelse(saison=="hiver", 1, 0))->deces_bailleul
deces_bailleul %>%
mutate(structures = "Bailleul")->deces_bailleul

```

#Fusion des tables

```

deces_bailleul %>%
rename(Commune = Communes)->deces_bailleul
UNA2019_2020 %>%
rbind(deces_bailleul)->UNA2019_2020

```

Lewarde

```

read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage évaluation impact aide
à domicile/données/études des données MAD/Arnaud stat/Arnaud
stat/deces_lewarde.csv", sep = ";", header = T)->deces_lewarde
deces_lewarde %>%
mutate(Date.de.décès = as.Date(Date.de.décès, format = "%d/%m/%Y")) %>%
mutate(panel = format(Date.de.décès, format = "%Y-%m")) %>%
filter(!is.na(age))->deces_lewarde
deces_lewarde %>%
select(-Date.de.décès)->deces_lewarde
deces_lewarde %>%
mutate(genre=ifelse(Sexe=="F",1,0)) %>%
mutate(confinement=ifelse(panel=="2020-03"|panel=="2020-04"|panel=="2020-
05",1,0)) %>%
mutate(GIR2 = as.factor(as.character(GIR2)))->deces_lewarde

```



```

deces_lewarde %>%
  full_join(revenu_communes, by = c("Commune" = "commune")) %>%
  filter(!is.na(panel))->deces_lewarde
deces_lewarde %>%
  inner_join(pop, by = c("Commune"="Commune"))->deces_lewarde
deces_lewarde$panel2[deces_lewarde$panel %in% c("2019-01", "2019-02", "2019-12", "2020-01", "2020-02")]<-"hiver"
deces_lewarde$panel2[deces_lewarde$panel %in% c("2019-03", "2019-04", "2019-05", "2020-03", "2020-04", "2020-05")]<-"printemps"
deces_lewarde$panel2[deces_lewarde$panel %in% c("2019-06", "2019-07", "2019-08")]<-"été"
deces_lewarde$panel2[deces_lewarde$panel %in% c("2019-09", "2019-10", "2019-11")]<-"automne"
deces_lewarde %>%
  mutate(saison = as.factor(panel2)) %>%
  select(-panel2)->deces_lewarde
deces_lewarde$categorie_commune<-cut(deces_lewarde$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes", "riches"))
deces_lewarde$classe_ville_pop<-cut(deces_lewarde$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))
deces_lewarde %>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
  mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
  mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
  mutate(automne = ifelse(saison=="automne", 1, 0)) %>%
  mutate(ete = ifelse(saison=="été", 1, 0)) %>%
  mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%
  mutate(hiver = ifelse(saison=="hiver", 1, 0))->deces_lewarde
deces_lewarde %>%
  mutate(structures = "LEWARDE")->deces_lewarde

```

#Fusion des tables

```

UNA2019_2020 %>%
  rbind(deces_lewarde)->UNA2019_2020

```

#ASSAD Dunkerque (supprimer de l'échantillon car données seulement disponible pendant le confinement)

```

assad_dunkerque<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/deces assad Dunkerque.csv", sep = ";", header = T, skip = 1)
assad_dunkerque %>%
  select(-c(X,X.1)) %>%
  filter(!is.na(age))->assad_dunkerque
assad_dunkerque %>%
  mutate(Date.de.décès = as.Date(Date.de.décès, format = "%d/%m/%Y")) %>%
  mutate(panel = format(Date.de.décès, format = "%Y-%m")) %>%
  select(-Date.de.décès)->assad_dunkerque

assad_dunkerque %>%
  mutate(genre=ifelse(Sexe=="F",1,0)) %>%
  mutate(confinement=ifelse(panel=="2020-03"|panel=="2020-04"|panel=="2020-
05",1,0)) %>%
  mutate(GIR2 = as.factor(as.character(GIR2)))->assad_dunkerque
assad_dunkerque %>%
  full_join(revenu_communes, by = c("Commune" = "commune")) %>%
  filter(!is.na(panel))->assad_dunkerque
assad_dunkerque %>%
  inner_join(pop, by = c("Commune"="Commune"))->assad_dunkerque
assad_dunkerque$panel2[assad_dunkerque$panel %in% c( "2019-01", "2019-02",
"2019-12", "2020-01", "2020-02")]<-"hiver"
assad_dunkerque$panel2[assad_dunkerque$panel %in% c("2019-03", "2019-04",
"2019-05", "2020-03", "2020-04", "2020-05")]<-"printemps"
assad_dunkerque$panel2[assad_dunkerque$panel %in% c("2019-06", "2019-07",
"2019-08")]<-"été"
assad_dunkerque$panel2[assad_dunkerque$panel %in% c("2019-09", "2019-10",
"2019-11")]<-"automne"
assad_dunkerque %>%
  mutate(saison = as.factor(panel2)) %>%
  select(-panel2)->assad_dunkerque
assad_dunkerque$categorie_commune<-cut(assad_dunkerque$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes","riches"))
assad_dunkerque$classe_ville_pop<-cut(assad_dunkerque$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))
assad_dunkerque %>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%

```

```

mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
mutate(automne = ifelse(saison=="automne", 1, 0)) %>%
mutate(ete = ifelse(saison=="été", 1, 0)) %>%
mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%
mutate(hiver = ifelse(saison=="hiver", 1, 0))->assad_dunkerque
assad_dunkerque %>%
mutate(structures = "ASSAD DUNKERQUE")->assad_dunkerque

```

#Fusion des tables

```

UNA2019_2020 %>%
  rbind(assad_dunkerque)->UNA2019_2020
#Suppression données ASSAD DUNKERQUE à cause d'un biais (seulement des
données pendant le confinement)
UNA2019_2020 %>%
  filter(structures!="ASSAD DUNKERQUE")->UNA2019_2020

```

#Gravelines

```

deces_gravelines<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/deces_gravelines.csv", sep = ";", header = T, skip = 1)
deces_gravelines %>%
  rename(GIR2=X.GIR2) %>%
  mutate(panel = format(Date.de.décès, format = "%Y-%m")) %>%
  filter(!is.na(age))->deces_gravelines
deces_gravelines %>%
  select(-Date.de.décès)->deces_gravelines

deces_gravelines %>%
  mutate(genre=ifelse(Sexe=="F",1,0)) %>%
  mutate(confinement=ifelse(panel=="2020-03"|panel=="2020-04"|panel=="2020-
05",1,0)) %>%
  mutate(GIR2 = as.factor(as.character(GIR2)))->deces_gravelines
deces_gravelines %>%
  full_join(revenu_communes, by = c("Commune" = "commune")) %>%
  filter(!is.na(panel))->deces_gravelines
deces_gravelines %>%
  inner_join(pop, by = c("Commune"="Commune"))->deces_gravelines
deces_gravelines$panel2[deces_gravelines$panel %in% c( "2019-01", "2019-02",
"2019-12", "2020-01", "2020-02")]<-"hiver"
deces_gravelines$panel2[deces_gravelines$panel %in% c("2019-03", "2019-04",
"2019-05", "2020-03", "2020-04", "2020-05")]<-"printemps"
deces_gravelines$panel2[deces_gravelines$panel %in% c("2019-06", "2019-07",
"2019-08")]<-"été"

```

```

deces_gravelines$panel2[deces_gravelines$panel %in% c("2019-09", "2019-10",
"2019-11")]<-"automne"
deces_gravelines %>%
  mutate(season = as.factor(panel2)) %>%
  select(-panel2)->deces_gravelines
deces_gravelines$categorie_commune<-cut(deces_gravelines$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes","riches"))
deces_gravelines$classe_ville_pop<-cut(deces_gravelines$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))
deces_gravelines %>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
  mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
  mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
  mutate(automne = ifelse(season=="automne", 1, 0)) %>%
  mutate(ete = ifelse(season=="été", 1, 0)) %>%
  mutate(printemps = ifelse(season=="printemps", 1, 0)) %>%
  mutate(hiver = ifelse(season=="hiver", 1, 0))->deces_gravelines
deces_gravelines %>%
  mutate(structures = "GRAVELINES")->deces_gravelines

```

#Fusion des tables

```

UNA2019_2020 %>%
  rbind(deces_gravelines)->UNA2019_2020

```

#STEENVOORDE

```

deces_steenvoorde<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/décès winnezele.csv", sep = ";", header = T, skip = 1)
deces_steenvoorde %>%
  mutate(panel = format(Date.de.décès, format = "%Y-%m")) %>%
  filter(!is.na(age))->deces_steenvoorde
deces_steenvoorde %>%
  select(-Date.de.décès)->deces_steenvoorde

deces_steenvoorde %>%
  mutate(genre=ifelse(Sexe=="F",1,0)) %>%
  mutate(confinement=ifelse(panel=="2020-03"|panel=="2020-04"|panel=="2020-
05",1,0)) %>%

```

```

rename(GIR2=X.GIR2) %>%
mutate(GIR2 = as.factor(as.character(GIR2)))->deces_steenvoorde
deces_steenvoorde %>%
  select(-c(X,X.1,X.2))->deces_steenvoorde
deces_steenvoorde %>%
  full_join(revenu_communes, by = c("Commune" = "commune")) %>%
  filter(!is.na(panel))->deces_steenvoorde
deces_steenvoorde %>%
  inner_join(pop, by = c("Commune"="Commune"))->deces_steenvoorde
deces_steenvoorde$panel2[deces_steenvoorde$panel %in% c( "2019-01", "2019-
02", "2019-12", "2020-01", "2020-02")]<-"hiver"
deces_steenvoorde$panel2[deces_steenvoorde$panel %in% c("2019-03", "2019-
04", "2019-05", "2020-03", "2020-04", "2020-05")]<-"printemps"
deces_steenvoorde$panel2[deces_steenvoorde$panel %in% c("2019-06", "2019-
07", "2019-08")]<-"été"
deces_steenvoorde$panel2[deces_steenvoorde$panel %in% c("2019-09", "2019-
10", "2019-11")]<-"automne"
deces_steenvoorde %>%
  mutate(saison = as.factor(panel2)) %>%
  select(-panel2)->deces_steenvoorde
deces_steenvoorde$categorie_commune<-cut(deces_steenvoorde$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes", "riches"))
deces_steenvoorde$classe_ville_pop<-cut(deces_steenvoorde$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))
deces_steenvoorde %>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
  mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
  mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
  mutate(automne = ifelse(saison=="automne", 1, 0)) %>%
  mutate(ete = ifelse(saison=="été", 1, 0)) %>%
  mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%
  mutate(hiver = ifelse(saison=="hiver", 1, 0))->deces_steenvoorde
deces_steenvoorde %>%
  mutate(structures = "STEENVOORDE")->deces_steenvoorde

#Fusion des tables
UNA2019_2020 %>%
  rbind(deces_steenvoorde)->UNA2019_2020

```

#Marcq en baroeul

```
deces_marcq<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/décès marcq en baroeul.csv", sep = ";", header = T, skip = 1)
deces_marcq %>%
  select(-c(X,X.1)) %>%
  filter(!is.na(age)) %>%
  rename(Commune=commune)->deces_marcq
deces_marcq %>%
  mutate(Date.de.décès = as.Date(Date.de.décès, format = "%d/%m/%Y")) %>%
  mutate(panel = format(Date.de.décès, format = "%Y-%m")) %>%
  select(-Date.de.décès)->deces_marcq

deces_marcq %>%
  mutate(genre=ifelse(Sexe=="F",1,0)) %>%
  mutate(confinement=ifelse(panel=="2020-03"|panel=="2020-04"|panel=="2020-
05",1,0)) %>%
  mutate(GIR2 = as.factor(as.character(GIR2)))->deces_marcq
deces_marcq %>%
  full_join(revenu_communes, by = c("Commune" = "commune")) %>%
  filter(!is.na(panel))->deces_marcq
deces_marcq %>%
  inner_join(pop, by = c("Commune"="Commune"))->deces_marcq
deces_marcq$panel2[deces_marcq$panel %in% c("2019-01", "2019-02", "2019-12",
"2020-01", "2020-02")]<-"hiver"
deces_marcq$panel2[deces_marcq$panel %in% c("2019-03", "2019-04", "2019-05",
"2020-03", "2020-04", "2020-05")]<-"printemps"
deces_marcq$panel2[deces_marcq$panel %in% c("2019-06", "2019-07", "2019-
08")]<-"été"
deces_marcq$panel2[deces_marcq$panel %in% c("2019-09", "2019-10", "2019-
11")]<-"automne"
deces_marcq %>%
  mutate(saison = as.factor(panel2)) %>%
  select(-panel2)->deces_marcq
deces_marcq$categorie_commune<-cut(deces_marcq$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes", "riches"))
deces_marcq$classe_ville_pop<-cut(deces_marcq$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))
deces_marcq %>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
```

```

mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
mutate(automne = ifelse(saison=="automne", 1, 0)) %>%
mutate(ete = ifelse(saison=="été", 1, 0)) %>%
mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%
mutate(hiver = ifelse(saison=="hiver", 1, 0))->deces_marcq
deces_marcq %>%
  mutate(structures = "MARCQ EN BAROEUL")->deces_marcq

```

#Fusion des tables

```

UNA2019_2020 %>%
  rbind(deces_marcq)->UNA2019_2020

```

#CAUDRY (supprimer de la base même biais que ASSAD DUNKERQUE)

```

CAUDRY<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage évaluation
impact aide à domicile/données/études des données MAD/Arnaud stat/Arnaud
stat/CAUDRY.csv", sep = ";", header = T, skip = 1)
CAUDRY %>%
  select(-c(X,X.1,X.2)) %>%
  filter(!is.na(age))->CAUDRY
CAUDRY%>%
  mutate(Date.de.décès = as.Date(Date.de.décès, format = "%d/%m/%Y")) %>%
  mutate(panel = format(Date.de.décès, format = "%Y-%m")) %>%
  select(-Date.de.décès)->CAUDRY

```

```

CAUDRY$Sexe[CAUDRY$Sexe %in% "FALSE"]<-"F"

```

```

CAUDRY %>%
  mutate(genre=ifelse(Sexe=="F",1,0)) %>%
  mutate(confinement=ifelse(panel=="2020-03"|panel=="2020-04"|panel=="2020-
05",1,0)) %>%
  mutate(GIR2 = as.factor(as.character(GIR2)))->CAUDRY
CAUDRY %>%
  full_join(revenu_communes, by = c("Commune" = "commune")) %>%
  filter(!is.na(panel))->CAUDRY
CAUDRY %>%
  inner_join(pop, by = c("Commune"="Commune"))->CAUDRY
CAUDRY$panel2[CAUDRY$panel %in% c( "2019-01", "2019-02", "2019-12", "2020-
01", "2020-02")]<-"hiver"
CAUDRY$panel2[CAUDRY$panel %in% c("2019-03", "2019-04", "2019-05", "2020-
03", "2020-04", "2020-05")]<-"printemps"

```

```

CAUDRY$panel2[CAUDRY$panel %in% c("2019-06", "2019-07", "2019-08")]<-"été"
CAUDRY$panel2[CAUDRY$panel %in% c("2019-09", "2019-10", "2019-11")]<-
"automne"
CAUDRY %>%
  mutate(saison = as.factor(panel2)) %>%
  select(-panel2)->CAUDRY
CAUDRY$categorie_commune<-cut(CAUDRY$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes","riches"))
CAUDRY$classe_ville_pop<-cut(CAUDRY$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))
CAUDRY %>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
  mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
  mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
  mutate(automne = ifelse(saison=="automne", 1, 0)) %>%
  mutate(ete = ifelse(saison=="été", 1, 0)) %>%
  mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%
  mutate(hiver = ifelse(saison=="hiver", 1, 0))->CAUDRY
CAUDRY %>%
  mutate(structures = "CAUDRY")->CAUDRY

```

#Fusion des tables

```

UNA2019_2020 %>%
  rbind(CAUDRY)->UNA2019_2020

```

#Suppression données CAUDRY à cause d'un biais (seulement des données pendant le confinement)

```

UNA2019_2020 %>%
  filter(structures!="CAUDRY")->UNA2019_2020

```

#SIN LE NOBLE

```

deces_sin<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage évaluation
impact aide à domicile/données/études des données MAD/Arnaud stat/Arnaud
stat/deces_SIN.csv", sep = ";", header = T, skip = 1)
deces_sin %>%
  select(-c(X,X.1, X.2)) %>%
  filter(!is.na(age))->deces_sin
deces_sin %>%

```



```

mutate(Date.de.décès = as.Date(Date.de.décès, format = "%d/%m/%Y")) %>%
mutate(panel = format(Date.de.décès, format = "%Y-%m")) %>%
select(-Date.de.décès)->deces_sin

deces_sin %>%
  mutate(genre=ifelse(Sexe=="F",1,0)) %>%
  mutate(confinement=ifelse(panel=="2020-03"|panel=="2020-04"|panel=="2020-
05",1,0)) %>%
  mutate(GIR2 = as.factor(as.character(GIR2)))->deces_sin
deces_sin %>%
  full_join(revenu_communes, by = c("Commune" = "commune")) %>%
  filter(!is.na(panel))->deces_sin
deces_sin %>%
  inner_join(pop, by = c("Commune"="Commune"))->deces_sin
deces_sin$panel2[deces_sin$panel %in% c("2019-01", "2019-02", "2019-12", "2020-
01", "2020-02")]<-"hiver"
deces_sin$panel2[deces_sin$panel %in% c("2019-03", "2019-04", "2019-05", "2020-
03", "2020-04", "2020-05")]<-"printemps"
deces_sin$panel2[deces_sin$panel %in% c("2019-06", "2019-07", "2019-08")]<-"été"
deces_sin$panel2[deces_sin$panel %in% c("2019-09", "2019-10", "2019-11")]<-"
automne"
deces_sin %>%
  mutate(saison = as.factor(panel2)) %>%
  select(-panel2)->deces_sin
deces_sin$categorie_commune<-cut(deces_sin$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes","riches"))
deces_sin$classe_ville_pop<-cut(deces_sin$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))
deces_sin%>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
  mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
  mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
  mutate(automne = ifelse(saison=="automne", 1, 0)) %>%
  mutate(ete = ifelse(saison=="été", 1, 0)) %>%
  mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%
  mutate(hiver = ifelse(saison=="hiver", 1, 0))->deces_sin
deces_sin %>%
  mutate(structures = "SIN LE NOBLE")->deces_sin

```

#Fusion des tables

```
UNA2019_2020 %>%  
  rbind(deces_sin)->UNA2019_2020
```

#DUNKERQUE

```
deces_DK<-read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage évaluation  
impact aide à domicile/données/études des données MAD/Arnaud stat/Arnaud  
stat/deces_dunkerque.csv", sep = ";", header = T, skip = 1)
```

```
deces_DK %>%  
  select(-c(X,X.1, X.2)) %>%  
  filter(!is.na(age))->deces_DK  
deces_DK %>%  
  mutate(Date.de.décès = as.Date(Date.de.décès, format = "%d/%m/%Y")) %>%  
  mutate(panel = format(Date.de.décès, format = "%Y-%m")) %>%  
  select(-Date.de.décès)->deces_DK
```

```
deces_DK %>%  
  mutate(genre=ifelse(Sexe=="F",1,0)) %>%  
  mutate(confinement=ifelse(panel=="2020-03"|panel=="2020-04"|panel=="2020-  
05",1,0)) %>%  
  mutate(GIR2 = as.factor(as.character(GIR2)))->deces_DK  
deces_DK %>%  
  full_join(revenu_communes, by = c("Commune" = "commune")) %>%  
  filter(!is.na(panel))->deces_DK  
deces_DK %>%  
  inner_join(pop, by = c("Commune"="Commune"))->deces_DK  
deces_DK$panel2[deces_DK$panel %in% c( "2019-01", "2019-02", "2019-12",  
"2020-01", "2020-02")]<-"hiver"  
deces_DK$panel2[deces_DK$panel %in% c("2019-03", "2019-04", "2019-05", "2020-  
03", "2020-04", "2020-05")]<-"printemps"  
deces_DK$panel2[deces_DK$panel %in% c("2019-06", "2019-07", "2019-08")]<-"  
été"  
deces_DK$panel2[deces_DK$panel %in% c("2019-09", "2019-10", "2019-11")]<-"  
automne"  
deces_DK %>%  
  mutate(saison = as.factor(panel2)) %>%  
  select(-panel2)->deces_DK  
deces_DK$categorie_commune<-cut(deces_DK$rev_median,  
c(0,19500,25700,35000), labels = c("pauvres", "moyennes", "riches"))  
deces_DK$classe_ville_pop<-cut(deces_DK$population,  
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",  
"grandes_villes"))
```

```

deces_DK %>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
  mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
  mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
  mutate(automne = ifelse(saison=="automne", 1, 0)) %>%
  mutate(ete = ifelse(saison=="été", 1, 0)) %>%
  mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%
  mutate(hiver = ifelse(saison=="hiver", 1, 0))->deces_DK
deces_DK %>%
  mutate(structures = "DUNKERQUE")->deces_DK

```

#Fusion des tables

```

UNA2019_2020 %>%
  rbind(deces_DK)->UNA2019_2020

```

Création des variables nombre de décès et log

```

UNA2019_2020 %>%
  group_by(panel) %>%
  mutate(nb.deces=n())->UNA2019_2020

UNA2019_2020 %>%
  mutate(log_age = log(age)) %>%
  mutate(log_deces = log(nb.deces))->UNA2019_2020

```

```

write.csv2(UNA2019_2020, file = "Stage-de-M2/base de
données/base_deces_Una2019_2020.csv")

```

#Base de données des hospitalisations

#hospitalisation MAD

```

read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage évaluation impact aide
à domicile/données/études des données MAD/Arnaud stat/Arnaud
stat/hospitalisation MAD 2019-2020.csv", header = T, sep = ";")->base_hospi
base_hospi %>%
  mutate(Debut.absence=as.Date(Debut.absence, format = "%d/%m/%Y")) %>%
  mutate(Fin.absence=as.Date(Fin.absence, format = "%d/%m/%Y")) %>%
  mutate(tps_hospi = trunc(time_length(interval(Debut.absence, Fin.absence),
"days"))->base_hospi
base_hospi %>%
  mutate(panel = format(Debut.absence, format = "%Y-%m")) %>%

```

```
filter(Debut.absence>="2019-01-01")->base_hospi
```

#Ajouter 1 au temps hospi à la fin

#Fusion base hospi avec données clients MAD

```
base_hospi %>%
```

```
  select(-c(Libelle.motif.absence,Libellé.hôpital,Libelle.motif.absence.1, Observations,  
Date.de.fin.d.hospitalisation,Date.de.début.d.hospitalisation))->base_hospi
```

```
base_hospi %>%
```

```
  mutate(Code.de.l.aide=as.character(Code.de.l.aide))->base_hospi
```

```
PanMAD %>%
```

```
  mutate(Code.Aidé = as.character(Code.Aidé))->PanMAD
```

```
base_hospi %>%
```

```
  full_join(PanMAD, by = c("Code.de.l.aide" = "Code.Aidé"))->base_hospi_MAD
```

```
base_hospi_MAD %>%
```

```
  filter(!is.na(panel)) %>%
```

```
  filter(!is.na(Commune))->base_hospi_MAD
```

```
base_hospi_MAD %>% mutate(date.fin = "31/05/2020")->base_hospi_MAD
```

```
base_hospi_MAD %>%mutate(date.fin = as.Date(date.fin, format = "%d/%m/%Y"))-  
>base_hospi_MAD
```

```
base_hospi_MAD %>%
```

```
  mutate(age2
```

=

```
  ifelse(!is.na(Date.décès),trunc(time_length(interval(Date.naissance,Date.décès),
```

```
"years")),      trunc(time_length(interval(Date.naissance,date.fin),      "years"))))-
```

```
>base_hospi_MAD
```

```
base_hospi_MAD %>%
```

```
  select(-c(Code.de.l.aide,      date.fin,      Date.décès,      Date.naissance,  
Situation.familiale,age)) %>%
```

```
  rename(age=age2)->base_hospi_MAD
```

```
base_hospi_MAD %>%
```

```
  rename(Sexe = Titre)->base_hospi_MAD
```

```
base_hospi_MAD %>%
```

```
  select(-Debut.absence,-Fin.absence)->base_hospi_MAD
```

#Création de la variable genre

```
base_hospi_MAD %>%
```

```
  mutate(genre=ifelse(Sexe=="F",1,0)) %>%
```

```
  select(-genres)->base_hospi_MAD
```

#Ajout des données de revenu et de population. Les données ont été retravaillées sur excel (Mise en majuscule des communes et suppression des accents, traits d'union et apostrophes)

```
base_hospi_MAD %>%
```

```
  full_join(revenu_communes, by = c("Commune" = "commune")) %>%
```

```
filter(!is.na(Sexe))->base_hospi_MAD
```

```
base_hospi_MAD %>%  
  full_join(pop, by = "Commune") %>%  
  filter(!is.na(Sexe))->base_hospi_MAD
```

```
base_hospi_MAD %>%  
  mutate(confinement = ifelse(panel=="2020-03"|panel=="2020-04"|panel=="2020-  
05", 1, 0))->base_hospi_MAD  
base_hospi_MAD$panel2[base_hospi_MAD$panel %in% c("2019-01", "2019-02",  
"2019-12", "2020-01", "2020-02")]<-"hiver"  
base_hospi_MAD$panel2[base_hospi_MAD$panel %in% c("2019-03", "2019-04",  
"2019-05", "2020-03", "2020-04", "2020-05")]<-"printemps"  
base_hospi_MAD$panel2[base_hospi_MAD$panel %in% c("2019-06", "2019-07",  
"2019-08")]<-"été"  
base_hospi_MAD$panel2[base_hospi_MAD$panel %in% c("2019-09", "2019-10",  
"2019-11")]<-"automne"  
base_hospi_MAD %>%  
  mutate(saison = as.factor(panel2)) %>%  
  select(-panel2)->base_hospi_MAD
```

#Construction des catégorie des communes

```
base_hospi_MAD$categorie_commune<-cut(base_hospi_MAD$rev_median,  
c(0,19500,25700,35000), labels = c("pauvres", "moyennes","riches"))  
base_hospi_MAD$classe_ville_pop<-cut(base_hospi_MAD$population,  
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",  
"grandes_villes"))
```

#Pour plus de facilité dans l'analyse descriptive les facteurs vont être transformé en numérique

```
base_hospi_MAD %>%  
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%  
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))  
%>%  
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%  
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%  
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%  
  mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%  
  mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%  
  mutate(automne = ifelse(saison=="automne", 1, 0)) %>%  
  mutate(ete = ifelse(saison=="été", 1, 0)) %>%  
  mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%  
  mutate(hiver = ifelse(saison=="hiver", 1, 0))->base_hospi_MAD
```

#Construction de la base hospi struct

```
base_hospi_MAD %>%  
  mutate(structures = "MAD")->base_hospi_struct
```

#Ajout des données disponibles pour les hospitalisations des structures adhérentes au réseau Una Nord. Suppression de Caudry, ASSAD Dunkerque et SIN LE NOBLE car les données d'hospitalisations ne sont disponibles que durant la période de confinement

```
hospi_bailleul=read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage  
évaluation impact aide à domicile/données/études des données MAD/Arnaud  
stat/Arnaud stat/hospitalisation bailleul.csv", sep = ";", header = T, skip = 1)
```

calcul du temps d'hospitalisation

```
hospi_bailleul %>%  
  mutate(Date.début.d.hospitalisation = as.Date(Date.début.d.hospitalisation, format =  
"%d/%m/%Y")) %>%  
  mutate(Date.fin.d.hospitalisation = as.Date(Date.fin.d.hospitalisation, format =  
"%d/%m/%Y")) %>%  
  mutate(tps_hospi = trunc(time_length(interval(Date.début.d.hospitalisation,  
Date.fin.d.hospitalisation), "days"))->hospi_bailleul  
hospi_bailleul %>%  
  filter(Date.début.d.hospitalisation>="2019-01-01" &  
Date.début.d.hospitalisation<="2020-05-31") %>%  
  mutate(panel = format(Date.début.d.hospitalisation, format = "%Y-%m")) %>%  
  select(-Date.début.d.hospitalisation, -Date.fin.d.hospitalisation)->hospi_bailleul
```

#Création des variables

```
hospi_bailleul %>%  
  mutate(genre = ifelse(Sexe=="F",1,0)) %>%  
  mutate(confinement = ifelse(panel=="2020-03" |panel=="2020-04"|panel=="2020-  
05", 1,0)) %>%  
  mutate(GIR2 = as.factor(as.character(GIR2)))->hospi_bailleul  
hospi_bailleul %>%  
  full_join(revenu_communes, by = c("Commune" = "commune")) %>%  
  filter(!is.na(panel))->hospi_bailleul  
hospi_bailleul %>%  
  inner_join(pop, by = c("Commune"="Commune"))->hospi_bailleul  
hospi_bailleul$panel2[hospi_bailleul$panel %in% c( "2019-01", "2019-02", "2019-12",  
"2020-01", "2020-02")]<-"hiver"  
hospi_bailleul$panel2[hospi_bailleul$panel %in% c("2019-03", "2019-04", "2019-05",  
"2020-03", "2020-04", "2020-05")]<-"printemps"
```

```

hospi_bailleul$panel2[hospi_bailleul$panel %in% c("2019-06", "2019-07", "2019-08")]<-"été"
hospi_bailleul$panel2[hospi_bailleul$panel %in% c("2019-09", "2019-10", "2019-11")]<-"automne"
hospi_bailleul %>%
  mutate(saison = as.factor(panel2)) %>%
  select(-panel2)->hospi_bailleul
hospi_bailleul$categorie_commune<-cut(hospi_bailleul$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes", "riches"))
hospi_bailleul$classe_ville_pop<-cut(hospi_bailleul$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))
hospi_bailleul %>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
  mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
  mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
  mutate(automne = ifelse(saison=="automne", 1, 0)) %>%
  mutate(ete = ifelse(saison=="été", 1, 0)) %>%
  mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%
  mutate(hiver = ifelse(saison=="hiver", 1, 0))->hospi_bailleul
hospi_bailleul %>%
  mutate(structures = "BAILLEUL")->hospi_bailleul

```

#Fusion des tables

```

base_hospi_struct %>%
  mutate(GIR2=as.factor(as.character(GIR2))) %>%
  rbind(hospi_bailleul)->base_hospi_struct

```

#LEWARDE

```

hospi_lewarde=read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/hospitalisation_Lewarde.csv", sep = ";", header = T)
# calcul du temps d'hospitalisation
hospi_lewarde %>%
  mutate(Date.début.d.hospitalisation = as.Date(Date.début.d.hospitalisation, format =
"%d/%m/%Y")) %>%
  mutate(Date.fin.d.hospitalisation = as.Date(Date.fin.d.hospitalisation, format =
"%d/%m/%Y")) %>%

```

```

mutate(tps_hospi = trunc(time_length(interval(Date.début.d.hospitalisation,
Date.fin.d.hospitalisation), "days"))->hospi_lewarde
hospi_lewarde %>%
  filter(Date.début.d.hospitalisation>="2019-01-01" &
Date.début.d.hospitalisation<="2020-05-31") %>%
  mutate(panel = format(Date.début.d.hospitalisation, format = "%Y-%m")) %>%
  select(-Date.début.d.hospitalisation, -Date.fin.d.hospitalisation)->hospi_lewarde
#Création des variables
hospi_lewarde %>%
  mutate(genre = ifelse(Sexe=="F",1,0)) %>%
  mutate(confinement = ifelse(panel=="2020-03" |panel=="2020-04"|panel=="2020-
05", 1,0)) %>%
  mutate(GIR2 = as.factor(as.character(GIR2)))->hospi_lewarde
hospi_lewarde %>%
  full_join(revenu_communes, by = c("Commune" = "commune")) %>%
  filter(!is.na(panel))->hospi_lewarde
hospi_lewarde %>%
  inner_join(pop, by = c("Commune"="Commune"))->hospi_lewarde
hospi_lewarde$panel2[hospi_lewarde$panel %in% c( "2019-01", "2019-02", "2019-
12", "2020-01", "2020-02")]<-"hiver"
hospi_lewarde$panel2[hospi_lewarde$panel %in% c("2019-03", "2019-04", "2019-
05", "2020-03", "2020-04", "2020-05")]<-"printemps"
hospi_lewarde$panel2[hospi_lewarde$panel %in% c("2019-06", "2019-07", "2019-
08")]<-"été"
hospi_lewarde$panel2[hospi_lewarde$panel %in% c("2019-09", "2019-10", "2019-
11")]<-"automne"
hospi_lewarde %>%
  mutate(saison = as.factor(panel2)) %>%
  select(-panel2)->hospi_lewarde
hospi_lewarde$categorie_commune<-cut(hospi_lewarde$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes","riches"))
hospi_lewarde$classe_ville_pop<-cut(hospi_lewarde$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))
hospi_lewarde%>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
  mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
  mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
  mutate(automne = ifelse(saison=="automne", 1, 0)) %>%

```



```

mutate(ete = ifelse(saison=="été", 1, 0)) %>%
mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%
mutate(hiver = ifelse(saison=="hiver", 1, 0))->hospi_lewarde
hospi_lewarde %>%
mutate(structures = "LEWARDE")->hospi_lewarde

```

#Fusion des tables

```

base_hospi_struct %>%
mutate(GIR2=as.factor(as.character(GIR2))) %>%
rbind(hospi_lewarde)->base_hospi_struct

```

#MARCQ EN BAROEUL

```

hospi_MEB=read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/hospitalisation MEB.csv", sep = ";", header = T, skip = 1)

```

calcul du temps d'hospitalisation

```

hospi_MEB %>%
mutate(Date.début.d.hospitalisation = as.Date(Date.début.d.hospitalisation, format =
"%d/%m/%Y")) %>%
mutate(Date.fin.d.hospitalisation = as.Date(Date.fin.d.hospitalisation, format =
"%d/%m/%Y")) %>%
mutate(tps_hospi = trunc(time_length(interval(Date.début.d.hospitalisation,
Date.fin.d.hospitalisation), "days"))->hospi_MEB
hospi_MEB %>%
filter(Date.début.d.hospitalisation>="2019-01-01" &
Date.début.d.hospitalisation<="2020-05-31") %>%
mutate(panel = format(Date.début.d.hospitalisation, format = "%Y-%m")) %>%
select(-Date.début.d.hospitalisation, -Date.fin.d.hospitalisation)->hospi_MEB
#Création des variables
hospi_MEB%>%
mutate(genre = ifelse(Sexe=="F",1,0)) %>%
mutate(confinement = ifelse(panel=="2020-03" |panel=="2020-04"|panel=="2020-
05", 1,0)) %>%
mutate(GIR2 = as.factor(as.character(GIR2)))->hospi_MEB
hospi_MEB %>%
full_join(revenu_communes, by = c("Commune" = "commune")) %>%
filter(!is.na(panel))->hospi_MEB
hospi_MEB %>%
inner_join(pop, by = c("Commune"="Commune"))->hospi_MEB
hospi_MEB$panel2[hospi_MEB$panel %in% c( "2019-01", "2019-02", "2019-12",
"2020-01", "2020-02")]<-"hiver"
hospi_MEB$panel2[hospi_MEB$panel %in% c("2019-03", "2019-04", "2019-05",
"2020-03", "2020-04", "2020-05")]<-"printemps"

```

```

hospi_MEB$panel2[hospi_MEB$panel %in% c("2019-06", "2019-07", "2019-08")]<-
"été"
hospi_MEB$panel2[hospi_MEB$panel %in% c("2019-09", "2019-10", "2019-11")]<-
"automne"
hospi_MEB %>%
  mutate(season = as.factor(panel2)) %>%
  select(-panel2)->hospi_MEB
hospi_MEB$categorie_commune<-cut(hospi_MEB$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes","riches"))
hospi_MEB$classe_ville_pop<-cut(hospi_MEB$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))
hospi_MEB%>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
  mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
  mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
  mutate(automne = ifelse(season=="automne", 1, 0)) %>%
  mutate(ete = ifelse(season=="été", 1, 0)) %>%
  mutate(printemps = ifelse(season=="printemps", 1, 0)) %>%
  mutate(hiver = ifelse(season=="hiver", 1, 0))->hospi_MEB
hospi_MEB %>%
  mutate(structures = "MARCQ EN BAROEUL")->hospi_MEB

```

#Fusion des tables

```

base_hospi_struct %>%
  mutate(GIR2=as.factor(as.character(GIR2))) %>%
  rbind(hospi_MEB)->base_hospi_struct

```

#STEENVOORDE

```

hospi_steenvoorde=read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/hospi steenvoorde.csv", sep = ";", header = T, skip = 1)
hospi_steenvoorde %>%
  select(-c(X, X.1,X.2)) %>%
  filter(Sexe!="")->hospi_steenvoorde

```

calcul du temps d'hospitalisation

```

hospi_steenvoorde %>%

```

```

mutate(Date.début.d.hospitalisation = as.Date(Date.début.d.hospitalisation, format =
"%d/%m/%Y")) %>%
mutate(Date.fin.d.hospitalisation = as.Date(Date.fin.d.hospitalisation, format =
"%d/%m/%Y")) %>%
mutate(tps_hospi = trunc(time_length(interval(Date.début.d.hospitalisation,
Date.fin.d.hospitalisation), "days"))->hospi_steenvoorde
hospi_steenvoorde %>%
filter(Date.début.d.hospitalisation>="2019-01-01" &
Date.début.d.hospitalisation<="2020-05-31") %>%
mutate(panel = format(Date.début.d.hospitalisation, format = "%Y-%m")) %>%
select(-Date.début.d.hospitalisation, -Date.fin.d.hospitalisation)->hospi_steenvoorde
#Création des variables
hospi_steenvoorde%>%
mutate(genre = ifelse(Sexe=="F",1,0)) %>%
mutate(confinement = ifelse(panel=="2020-03" |panel=="2020-04"|panel=="2020-
05", 1,0)) %>%
mutate(GIR2 = as.factor(as.character(GIR2)))->hospi_steenvoorde
hospi_steenvoorde %>%
full_join(revenu_communes, by = c("Commune" = "commune")) %>%
filter(!is.na(panel))->hospi_steenvoorde
hospi_steenvoorde %>%
inner_join(pop, by = c("Commune"="Commune"))->hospi_steenvoorde
hospi_steenvoorde$panel2[hospi_steenvoorde$panel %in% c( "2019-01", "2019-02",
"2019-12", "2020-01", "2020-02")]<-"hiver"
hospi_steenvoorde$panel2[hospi_steenvoorde$panel %in% c("2019-03", "2019-04",
"2019-05", "2020-03", "2020-04", "2020-05")]<-"printemps"
hospi_steenvoorde$panel2[hospi_steenvoorde$panel %in% c("2019-06", "2019-07",
"2019-08")]<-"été"
hospi_steenvoorde$panel2[hospi_steenvoorde$panel %in% c("2019-09", "2019-10",
"2019-11")]<-"automne"
hospi_steenvoorde %>%
mutate(saison = as.factor(panel2)) %>%
select(-panel2)->hospi_steenvoorde
hospi_steenvoorde$categorie_commune<-cut(hospi_steenvoorde$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes", "riches"))
hospi_steenvoorde$classe_ville_pop<-cut(hospi_steenvoorde$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))
hospi_steenvoorde%>%
mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%

```

```

mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
mutate(automne = ifelse(saison=="automne", 1, 0)) %>%
mutate(ete = ifelse(saison=="été", 1, 0)) %>%
mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%
mutate(hiver = ifelse(saison=="hiver", 1, 0))->hospi_steenvoorde
hospi_steenvoorde %>%
mutate(structures = "STEENVOORDE")->hospi_steenvoorde

```

#Fusion des tables

```

base_hospi_struct %>%
mutate(GIR2=as.factor(as.character(GIR2))) %>%
rbind(hospi_steenvoorde)->base_hospi_struct

```

#DUNKERQUE

```

hospi_DK=read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage évaluation
impact aide à domicile/données/études des données MAD/Arnaud stat/Arnaud
stat/hospi DK.csv", sep = ";", header = T, skip = 1)
hospi_DK %>%
select(-c(X, X.1,X.2)) %>%
filter(Sexe!="")->hospi_DK

```

calcul du temps d'hospitalisation

```

hospi_DK %>%
mutate(Date.debut.d.hospitalisation = as.Date(Date.debut.d.hospitalisation, format =
"%d/%m/%Y")) %>%
mutate(Date.fin.d.hospitalisation = as.Date(Date.fin.d.hospitalisation, format =
"%d/%m/%Y")) %>%
mutate(tps_hospi = trunc(time_length(interval(Date.debut.d.hospitalisation,
Date.fin.d.hospitalisation), "days"))->hospi_DK
hospi_DK %>%
filter(Date.debut.d.hospitalisation>="2019-01-01" &
Date.debut.d.hospitalisation<="2020-05-31") %>%
mutate(panel = format(Date.debut.d.hospitalisation, format = "%Y-%m")) %>%
select(-Date.debut.d.hospitalisation, -Date.fin.d.hospitalisation)->hospi_DK

```

#Création des variables

```

hospi_DK%>%
mutate(genre = ifelse(Sexe=="F",1,0)) %>%
mutate(confinement = ifelse(panel=="2020-03" |panel=="2020-04"|panel=="2020-
05", 1,0)) %>%
mutate(GIR2 = as.factor(as.character(GIR2)))->hospi_DK
hospi_DK %>%

```

```

full_join(revenu_communes, by = c("Commune" = "commune")) %>%
  filter(!is.na(panel))->hospi_DK
hospi_DK %>%
  inner_join(pop, by = c("Commune"="Commune"))->hospi_DK
hospi_DK$panel2[hospi_DK$panel %in% c("2019-01", "2019-02", "2019-12", "2020-01", "2020-02")]<-"hiver"
hospi_DK$panel2[hospi_DK$panel %in% c("2019-03", "2019-04", "2019-05", "2020-03", "2020-04", "2020-05")]<-"printemps"
hospi_DK$panel2[hospi_DK$panel %in% c("2019-06", "2019-07", "2019-08")]<-"été"
hospi_DK$panel2[hospi_DK$panel %in% c("2019-09", "2019-10", "2019-11")]<-"automne"
hospi_DK %>%
  mutate(season = as.factor(panel2)) %>%
  select(-panel2)->hospi_DK
hospi_DK$categorie_commune<-cut(hospi_DK$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes", "riches"))
hospi_DK$classe_ville_pop<-cut(hospi_DK$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes", "grandes_villes"))
hospi_DK%>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
  mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
  mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
  mutate(automne = ifelse(season=="automne", 1, 0)) %>%
  mutate(ete = ifelse(season=="été", 1, 0)) %>%
  mutate(printemps = ifelse(season=="printemps", 1, 0)) %>%
  mutate(hiver = ifelse(season=="hiver", 1, 0))->hospi_DK
hospi_DK %>%
  mutate(structures = "DUNKERQUE")->hospi_DK

```

#Fusion des tables

```

base_hospi_struct %>%
  mutate(GIR2=as.factor(as.character(GIR2))) %>%
  rbind(hospi_DK)->base_hospi_struct

```

#GRAVELINES

```

hospi_gravelines=read.csv2(file = "C:/Users/jupiter/Desktop/Arnaud Blanc/stage
évaluation impact aide à domicile/données/études des données MAD/Arnaud
stat/Arnaud stat/hospi_gravelines.csv", sep = ";", header = T, skip = 1)

```

```
hospi_gravelines %>%
  select(-c(X, X.1)) %>%
  filter(Sexe!="")->hospi_gravelines
```

calcul du temps d'hospitalisation

```
hospi_gravelines %>%
  mutate(Date.début.d.hospitalisation = as.Date(Date.début.d.hospitalisation, format =
"%d/%m/%Y")) %>%
  mutate(Date.fin.d.hospitalisation = as.Date(Date.fin.d.hospitalisation, format =
"%d/%m/%Y")) %>%
  mutate(tps_hospi = trunc(time_length(interval(Date.début.d.hospitalisation,
Date.fin.d.hospitalisation), "days"))->hospi_gravelines
hospi_gravelines %>%
  filter(Date.début.d.hospitalisation>="2019-01-01" &
Date.début.d.hospitalisation<="2020-05-31") %>%
  mutate(panel = format(Date.début.d.hospitalisation, format = "%Y-%m")) %>%
  select(-Date.début.d.hospitalisation, -Date.fin.d.hospitalisation)->hospi_gravelines
```

#Création des variables

```
hospi_gravelines%>%
  mutate(genre = ifelse(Sexe=="F",1,0)) %>%
  mutate(confinement = ifelse(panel=="2020-03" |panel=="2020-04"|panel=="2020-
05", 1,0)) %>%
  mutate(GIR2 = as.factor(as.character(GIR2)))->hospi_gravelines
hospi_gravelines %>%
  full_join(revenu_communes, by = c("Commune" = "commune")) %>%
  filter(!is.na(panel))->hospi_gravelines
hospi_gravelines %>%
  inner_join(pop, by = c("Commune"="Commune"))->hospi_gravelines
hospi_gravelines$panel2[hospi_gravelines$panel %in% c( "2019-01", "2019-02",
"2019-12", "2020-01", "2020-02")]<-"hiver"
hospi_gravelines$panel2[hospi_gravelines$panel %in% c("2019-03", "2019-04",
"2019-05", "2020-03", "2020-04", "2020-05")]<-"printemps"
hospi_gravelines$panel2[hospi_gravelines$panel %in% c("2019-06", "2019-07",
"2019-08")]<-"été"
hospi_gravelines$panel2[hospi_gravelines$panel %in% c("2019-09", "2019-10",
"2019-11")]<-"automne"
hospi_gravelines %>%
  mutate(saison = as.factor(panel2)) %>%
  select(-panel2)->hospi_gravelines
hospi_gravelines$categorie_commune<-cut(hospi_gravelines$rev_median,
c(0,19500,25700,35000), labels = c("pauvres", "moyennes", "riches"))
```

```

hospi_gravelines$classe_ville_pop<-cut(hospi_gravelines$population,
c(0,5000,20000,100000,235000), labels = c("villages", "petites_villes", "villes",
"grandes_villes"))
hospi_gravelines%>%
  mutate(communes_riche = ifelse(categorie_commune=="riches",1,0)) %>%
  mutate(communes_moyennes = ifelse(categorie_commune=="moyennes",1,0))
%>%
  mutate(communes_pauvres = ifelse(categorie_commune=="pauvres", 1,0)) %>%
  mutate(villages = ifelse(classe_ville_pop=="villages", 1, 0)) %>%
  mutate(petites_villes = ifelse(classe_ville_pop=="petites_villes", 1, 0)) %>%
  mutate(villes = ifelse(classe_ville_pop=="villes", 1, 0)) %>%
  mutate(grandes_villes = ifelse(classe_ville_pop=="grandes_villes", 1, 0)) %>%
  mutate(automne = ifelse(saison=="automne", 1, 0)) %>%
  mutate(ete = ifelse(saison=="été", 1, 0)) %>%
  mutate(printemps = ifelse(saison=="printemps", 1, 0)) %>%
  mutate(hiver = ifelse(saison=="hiver", 1, 0))->hospi_gravelines
hospi_gravelines %>%
  mutate(structures = "GRAVELINES")->hospi_gravelines
#Fusion des tables
base_hospi_struct %>%
  mutate(GIR2=as.factor(as.character(GIR2))) %>%
  rbind(hospi_gravelines)->base_hospi_struct

#Finition sur la base des hospitalisations
base_hospi_struct %>%
  group_by(panel) %>%
  mutate(nb_hospitalisation = n()) %>%
  mutate(tps_hospi = tps_hospi+1) %>%
  mutate(log_age = log(age)) %>%
  mutate(log_tps = log(tps_hospi)) %>%
  mutate(log_hospi = log(nb_hospitalisation))->base_hospi_struct

write.csv2(base_hospi_struct, file = "Stage-de-M2/base de
données/base_hospi_una2019_2020.csv")

```

L'impact social du covid-19 : Décès et hospitalisations dans les services d'aide à domicile

L'Union Nationales de l'Aide, des Soins et des Services aux Domiciles
du département du Nord

Depuis le mois de février, la région des Hauts-de-France est touchée par une épidémie mondiale, le coronavirus. Cette pandémie a provoqué une surmortalité. Celle-ci provoque des décès en majorité chez les personnes âgées. Or, ces personnes sont en majorité prises en charge par les services d'aide à domicile. En effet, les associations d'aide à domicile qui sont encore dominantes sur ce marché prennent majoritairement en charge des personnes âgées dépendantes. Cette étude sur l'impact de la covid-19 repose sur deux indicateurs clés : les décès et les hospitalisations. Pour mesurer cet impact, une approche économétrique a été mobilisée. Ainsi, la méthode des moindres carrés ordinaires a été utilisée. Malheureusement, cette méthode fait face à un biais important. Pour corriger ce défaut, les estimations seront corrigées par l'utilisation des moindres carrés généralisés. Ainsi, la période de confinement a permis de réduire les prises en charge à l'hôpital et donc de générer des économies importantes.

Mots clefs : coronavirus, décès et hospitalisations, personnes âgées dépendantes, confinement, économies