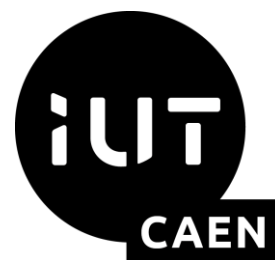




UNIVERSITÉ
CAEN
NORMANDIE



UNIVERSITÉ DE CAEN NORMANDIE

Institut Universitaire de Technologie de CAEN

Département Statistique et Informatique Décisionnelle

Diplôme Universitaire de Technologie
STATISTIQUE ET INFORMATIQUE DÉCISIONNELLE

Exploitation de bases de données relatives à la mise en réseaux de structures hospitalières

Rapport de Stage



Hugo MAURY

Maître de stage : LE MEUR Nolwenn
Tuteur de stage : EL HASSAN Fadili

Année universitaire 2021/2022
Stage effectué du 04/04/2022 au 17/06/2022

Exploitation de bases de données
relatives à la mise en réseaux de
structures hospitalières

1 Remerciement

Je remercie mon maître de stage, Nolwenn Le Meur, de m'avoir accueilli dans son service et fait découvrir le monde de la santé publique. Je l'a remercie aussi de m'avoir dirigé et permis de profiter pleinement de ces 11 semaines de stage.

Je remercie mon tuteur de stage, Fadili El Hassan, pour m'avoir soutenu durant ce stage.

Je remercie également toute l'équipe pour leur accueil, leur sympathie et leur présence tout au long du stage.

2 Résumé de l'étude

En Français

L'EHESP et l'équipe de recherche RSMS¹, au sein de laquelle j'ai réalisé mon stage, est partie prenante de l'évaluation d'une expérimentation nationale de la rémunération des hôpitaux (et de leurs partenaires) à l'épisode de soins au lieu du paiement à l'acte.

L'objectif de ce stage est de comparer le profil des hôpitaux volontaires au projet, à ceux non-volontaire, ainsi que déterminer le profil type de l'hôpital, dont le forfait serait le plus avantageux.

Les données proviennent du programme de médicalisation des systèmes d'information, qui regroupe les données de l'ensemble des patients hospitalisés, de 2015 à 2019, ainsi que du site atlas s@nté.

Ainsi, selon nos résultats, il semblerait avoir des différences entre ces parties, au niveau de leurs réseaux ainsi que leurs offres de soins.

En Anglais

The EHESP and the RSMS research team, in which I did my internship, is involved in the evaluation of a national experiment of the compensation of hospitals (and their partners) at the episode of care instead of the fee-for-service.

The objective of this internship is to compare the profile of the volunteer hospitals to the project, as well as to determine the typical profile of the hospital, whose package would be the most advantageous.

The data comes from the Medicalization of Information Systems program, which aggregates data from all hospitalized patients from 2015 to 2019, as well as the atlas s@nté site.

Thus, according to our results, there appear to be differences between these parties, in terms of their networks as well as their healthcare offerings.

¹ Recherche sur les Services et le Management en Santé, source 1

SOMMAIRE

1	Remerciement	4
2	Résumé de l'étude.....	5
3	Présentation de l'Ecole des Hautes Etudes en Santé Publique (EHESP).....	8
3.1	Point historique	8
3.2	Les départements au sein de l'école	9
3.3	METIS et la recherche	10
4	Contexte du stage	12
4.1	Le projet Paiement à l'Episode de Soin (EDS)	12
4.2	Objectif et déroulement du stage	14
5	Analyse de la structure des réseaux hospitaliers et leurs impacts dans l'EDS	16
5.1	Matériels	16
5.2	Méthodes	18
5.3	Résultats	19
5.4	Conclusion	26
6	Analyse des indicateurs territoriaux des établissements	27
6.1	Matériels	27
6.2	Méthodes	29
6.3	Résultats	30
6.4	Conclusion	37
7	Conclusion	38
8	Bilan personnel	38
9	Référence	39
10	Annexe	40
10.1	Programmation	40

Table des figures

<i>Figure 1 : Organigramme de l'EHESP.....</i>	<i>9</i>
<i>Figure 2 : Carte de France situant les hôpitaux participants à l'expérimentation du projet EDS.....</i>	<i>12</i>
<i>Figure 3 : Capture d'écran de la trame de communication pour le 58ème colloques</i>	<i>15</i>
<i>Figure 4 : Matrice de corrélation des variables.....</i>	<i>21</i>
<i>Figure 5 : ANOVA du modèle logistique complet</i>	<i>22</i>
<i>Figure 6 : Modèle logistique final</i>	<i>22</i>
<i>Figure 7 : Courbe de sensibilité et de spécificité</i>	<i>24</i>
<i>Figure 8 : Graphe représentant les réseaux de transfère des patients, venant des hôpitaux expérimentateurs.....</i>	<i>25</i>
<i>Figure 9 : Distribution de la densité de kinésithérapeutes, pour 10 000 habitants, à partir des simulations des établissements gagnants/perdants, en 2019.....</i>	<i>30</i>
<i>Tableau 1 : Planning du stage.....</i>	<i>14</i>
<i>Tableau 2 : Extrait du fichier de données.....</i>	<i>18</i>
<i>Tableau 3 : Comparaison, année par année, des indicateurs réseaux des hôpitaux expérimentateurs et non-expérimentateurs.....</i>	<i>19</i>
<i>Tableau 4 : Comparaison des indicateurs réseaux, sur la période 2015 – 2019, des hôpitaux expérimentateurs et non-expérimentateurs.....</i>	<i>20</i>
<i>Tableau 5 : Matrice de confusion</i>	<i>24</i>
<i>Tableau 6 : Extrait du fichier de données.....</i>	<i>29</i>
<i>Tableau 7 : Comparaison des indicateurs territoriaux, à partir de la simulation des établissements de soins estimés gagnants et perdants.....</i>	<i>31</i>
<i>Tableau 8 : Comparaison des indicateurs territoriaux, concernant le type de soin sur la prothèse totale de hanche, à partir des établissements estimés gagnants et perdants</i>	<i>33</i>
<i>Tableau 9 : Comparaison des indicateurs territoriaux, concernant le type de soin sur la prothèse totale de genou, à partir des établissements estimés gagnants et perdants.....</i>	<i>34</i>
<i>Tableau 10 : Odds Ratio de la régression logistique multivariée, à partir de la simulation des établissements de soins estimés gagnants et perdants, au niveau global.....</i>	<i>35</i>
<i>Tableau 11 : Coefficient de l'analyse en panel de la complexité des patients venant d'établissements MCO, vers des établissements SSR, pour les établissements de soins estimés gagnants.....</i>	<i>36</i>
<i>Tableau 12 : Coefficient de l'analyse en panel du taux d'adressage (%) de patient venant d'établissements MCO, vers des établissements SSR, pour les établissements de soins estimés gagnants.....</i>	<i>37</i>

3 Présentation de l'Ecole des Hautes Etudes en Santé Publique (EHESP)

Pour mon stage de fin d'année, j'ai eu la chance de pouvoir être accueillie par un organisme reconnu nationalement, une école avec un service développé concernant les statistiques et l'informatique, l'Ecole des Hautes Etudes en Santé Publique (EHESP). Avant de commencer, je vais vous présenter l'organisme. Dans un premier temps, je vais vous présenter son histoire². Par la suite, je ferai une présentation des principaux départements de l'école³. Enfin, je terminerai une présentation du pôle dans lequel j'effectue mon stage.

3.1 Point historique

Cette école a été fondée en 1945, sous le nom de l'École Nationale de la Santé Publique (ENSP), durant la 4ème république, avec comme objectif la mise en place de grandes réformes portant sur la protection sociale. C'est pour cela que, sous l'égide du ministère de la Santé et des Affaires Sociales, l'école forme des cadres supérieurs qui pourront appliquer ces nouvelles politiques. Au tout début, l'ENSP dépend principalement de l'institut national d'hygiène, qui avait pour but de gérer aussi bien le travail des médecins, que des pharmaciens, sur des sujets médicaux et administratifs. Elle sera obligée de migrer vers Rennes en 1962, dû à un manque de personnel enseignant permanent, face à la demande de plus en plus pressante en matière d'éradication des épidémies. Mais grâce à ce transfert, l'école dispose désormais d'un véritable campus, ainsi que d'un budget plus convenable. Le premier directeur de l'établissement de Rennes sera le professeur de pédiatrie Jean Senecal. Elle continuera à suivre les volontés des créateurs en formant des cadres supérieurs centrés sur le secteur sanitaire et social, mais une volonté d'étendre le champ d'action de la santé publique au sens international fera pousser l'idée de faire évoluer l'ENSP vers une nouvelle structure. Ce fut le directeur et professeur Jacques Hardy, qui sera chargé de préparer cette évolution vers 2006. L'Ecole des Hautes Etudes en Santé Publique voit le jour le 1er janvier 2008, sous deux tutelles, le ministère de la santé et le ministère de l'Education nationale, avec comme objectif la poursuite de la mission de l'ENSP, assurer la formation statutaire de nombreux métiers du sanitaire et du social et aura comme second objectif le développement de la recherche et l'élargissement à l'international. Tout cela sera supervisé par Antoine Flahault, premier directeur de

² Source 2 : Présentation historique de l'EHESP

³ Source 3 : Présentation de l'EHESP

l'Ecole des Hautes en Santé Publique et praticien hospitalier. Par la suite, c'est Laurent Chambaud, qui prendra sa place depuis 2013.

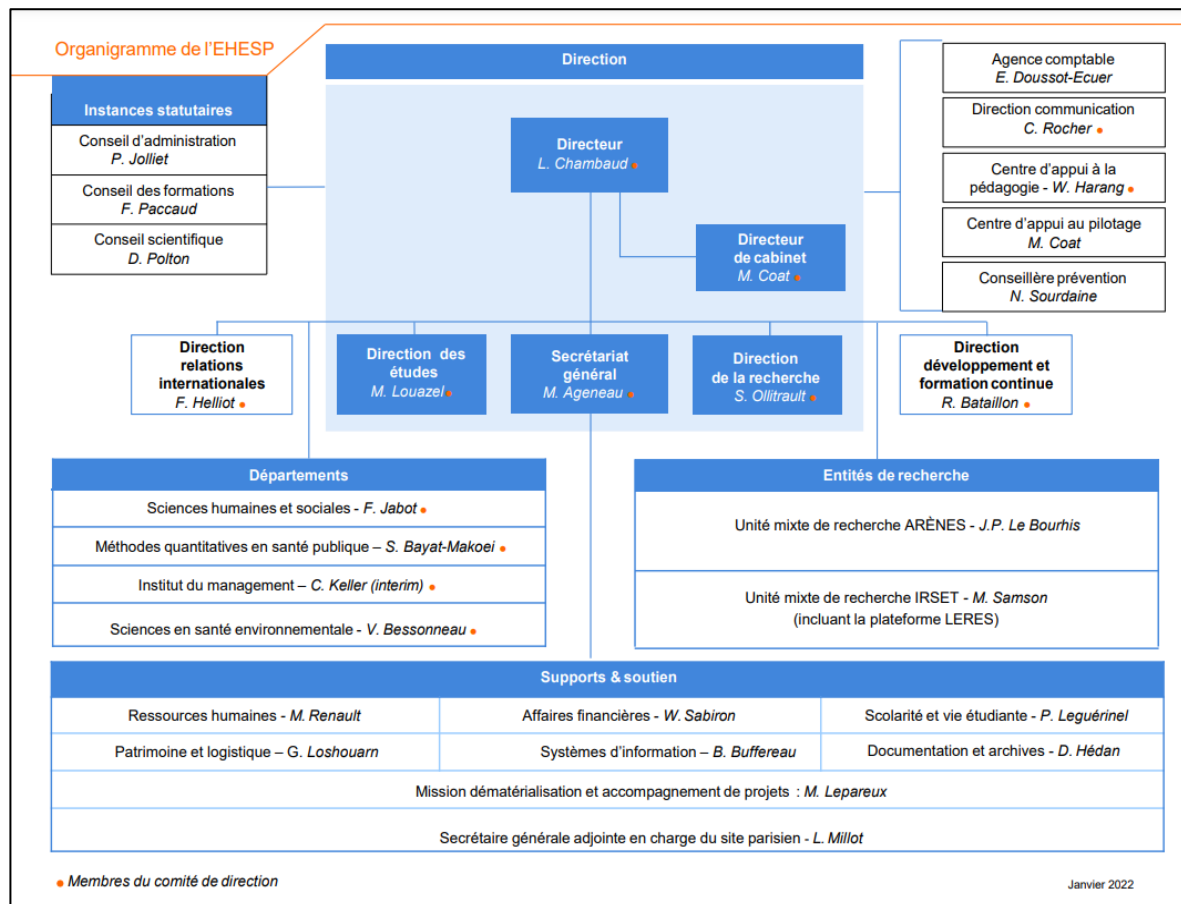


Figure 1 : Organigramme de l'EHESP

3.2 Les départements au sein de l'école

L'école est située à Rennes, avec une antenne à Paris, et est composée de plus de 400 personnes, dont 90 professeurs qui sont regroupés dans 4 départements d'enseignement et de recherche (*figure 1*). Nous retrouvons le département méthodes quantitatives en santé publique (METIS), dans lequel j'effectue mon stage, le département science en santé environnementale (SSE), le département Institut du management (IDM) et pour finir le département des sciences humaines et sociales (SHS).

Le département IDM (*figure 1*), a pour mission de développer les compétences et l'expertise en management des établissements et des services, mais aussi des politiques de santé. L'institut du management intervient dans le domaine de la stratégie, de l'analyse et dans la gouvernance des systèmes et des organisations, ou dans la gestion des opérateurs et des technologies, pour ne citer que des exemples.

Concernant le département SHS (*figure 1*), ces objectifs principaux sont portés sur les politiques publiques développées dans le champ sanitaire et social à partir des données des sciences humaines et sociales. Pour répondre à ces objectifs, ils se penchent sur la vulnérabilité et les liens sociaux, les déterminants sociaux et la promotion de la santé, ou encore l'analyse des politiques sociales et de la santé.

Le département SSE (*figure 1*), dispense des formations pour les cadres et opérateurs du système de sécurité sanitaire environnementales aux plans nationaux, et le développement de recherche compétitive au plan international sur les dangers et les risques liés aux milieux de vie et du travail, ainsi que leur prévention en intégrant les différents champs disciplinaires pertinents. Leurs missions sont principalement orientées sur les risques pour la santé liée aux milieux de vie.

Pour METIS (*figure 1*), les actions portent sur les activités de formation et de recherche en épidémiologie, la statistiques et systèmes d'information pour la santé des populations destinées à guider l'action en santé publique. Cette équipe est sous la responsabilité de Sahar Bayat, depuis le 1^{er} janvier 2022. Leur principale mission est orientée sur la sécurité sanitaire ou encore la prévention et la promotion de la santé.

Mais en dehors des différents pôles de recherche, nous retrouvons, avant tout, une école avec comme but principal la formation. Pour intégrer la formation fonction publique de cette école, il faut nécessairement être titulaire d'un diplôme (minimum bac + 3). De plus, elle est accessible uniquement sur concours. Les concours visent la formation initiale et le perfectionnement de fonctionnaires de catégorie A. Elle propose aussi un cursus aux préparations aux concours, aussi bien sur les épreuves écrites, que orales. Cette formation est la meilleure manière de pouvoir avoir une chance d'intégrer la formation fonction publique.

3.3 METIS et la recherche

Je vais faire une présentation plus approfondie du département, au sein du quel j'effectuerai ma mission. METIS (Méthodes quantitatives en santé publique) est un département qui mène des recherches s'inscrivant dans les grandes thématiques de l'école.

Ce département a une forte valence de recherche, puisqu'elle est composée uniquement d'enseignants-chercheurs. Concernant leur mission, elle est orientée sur la sécurité sanitaire, analysant les risques sanitaires au travail, la sécurité des soins, la sécurité des produits de santé, la modélisation des maladies émergentes grâce à des indicateurs statistiques. Elle fait également de la promotion et de la prévention consacrée à la santé mère-enfant et aux programmes de vaccination. Leurs recherches touchent aussi à l'environnement, avec des analyses sur la contribution des expositions

environnementales aux inégalités sociales de santé. Et pour finir, nous pouvons retrouver des chercheurs se penchant sur la question de l'organisation et le management des services de santé, avec comme préoccupation la prise en charge des maladies chroniques (neurologiques, vasculaires, rénales).

Pour mener à bien ces recherches, le département s'appuie sur des méthodes quantitatives. Le département a les compétences nécessaires pour l'exploitation des bases de données médico-sociales, sanitaires et administratives : Programme de Médicalisation des Systèmes d'Information (PMSI), Système national d'information inter-régimes de l'assurance maladie (SNIIRAM) ...

Les enseignements dispensés en épidémiologie et en bio statistiques par le département, ont pour but de transmettre aux futurs professionnels de santé, les bases pour la compréhension des principaux outils et des indicateurs utilisés en santé publique, pour la mesure de l'état de santé de la population et de l'organisation de l'offre de soin. Des intervenants extérieurs, experts nationaux et internationaux prennent régulièrement part aux formations.

Le département assure la coordination du socle commun en santé publique, dans le cadre du séminaire de rentrée des filières fonctions publiques, et, organise les modules en sciences quantitatives en santé populationnelle dispensés à ces filières professionnelles. Des cours sont délivrés en anglais, dans le cadre du Master de Santé Publique (MPH) et d'Europubhealth (European Public Health Master), dont le département porte plusieurs parcours. En formation continue, METIS coordonne des sessions sur les thèmes du Programme de Médicalisation des Systèmes d'Information (PMSI), et de la sécurité sanitaire et gestion des risques. Il délivre un certificat international d'épidémiologie appliquée (IDEA) et un diplôme d'établissement Expert PMSI. Il est responsable de la formation des Pharmaciens Inspecteurs de Santé Publique (PHISP), en coordonnant la formation initiale, en organisant le stage statutaire (séminaire annuel regroupant l'ensemble du corps des PHISP), et en proposant une offre de formation continue adaptée à leurs besoins et activités. Le département coordonne et enseigne, avec l'Université de Rennes 1 et Rennes 2, le Master de Santé Publique (parcours METEORES), le Master 1 Sciences Sociales et management, le Master 1 santé publique parcours sciences quantitatives, et est partenaire du Master spécialisé en santé publique de l'École Pasteur/CNAM à Paris.

4 Contexte du stage

4.1 Le projet Paiement à l'Episode de Soin ⁴(EDS)

Ce stage s'inscrit au projet Episode de soin⁵, visant à tester, pendant une durée de 5 ans auprès d'acteurs volontaires, la mise en place d'un modèle prévoyant un paiement forfaitaire ciblé pour l'ensemble des prestations définies dans un épisode de soins donné (les prises en charge PTH⁶, PTG⁷ et la colectomie pour cancer sont ciblées à ce jour). Elle poursuit un objectif d'amélioration de la qualité et de la sécurité des prises en charge, de l'efficacité des soins et de la satisfaction des patients, grâce à une organisation intégrée fondée sur les bonnes pratiques de prise en charge et la coordination des acteurs intra et extrahospitaliers. Ce projet a été lancé le 17 juillet 2019 par le Ministère des Solidarités et de la santé et l'Assurance Maladie, avec l'aide de l'Article 51, dont l'objectif est de favoriser l'innovation en santé, notamment le parcours de soin.

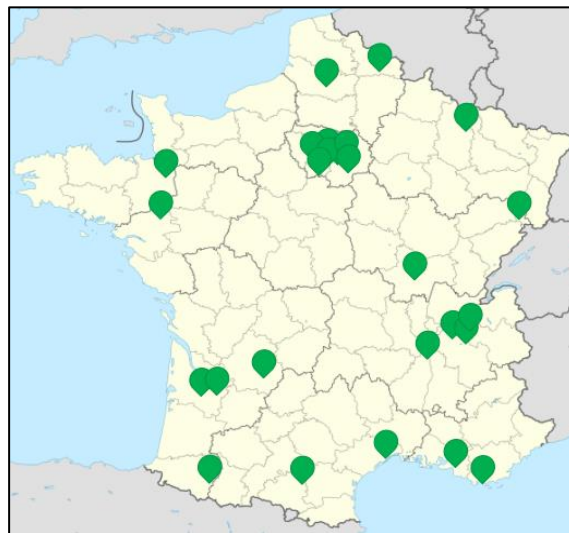


Figure 2 : Carte de France situant les hôpitaux participants à l'expérimentation du projet EDS

Le forfait comprendra toutes les dépenses comprises dans le périmètre défini de l'épisode de soins, ville et hôpital, pré, per et post-opératoire. Il couvrira les ressources nécessaires à la coordination, intégrera le risque de réhospitalisation et prendra également en compte la qualité via un compartiment dédié. Il est ajusté pour chaque patient, en prenant en compte leurs facteurs de risques et certaines caractéristiques du séjour initial. Au niveau nationale, par secteur, il est calculé sur une base historique à partir des dépenses encadrées par l'Assurance Maladie, et correspondants aux dépenses moyennes observées.

Il y a 25 candidats (*figure 2*, 29 établissements de santé), porteurs de projet de tous statuts et leurs partenaires participent volontairement à l'expérimentation, dont 24 concernant la colectomie pour cancer et 29 concernant les PTH et PTG.

⁴ Source 4: Présentation du projet Episode De Soins

⁵ L'épisode de soins représente l'ensemble des prestations dispensées à un patient pour un état de santé donné, pendant une période de temps définie, par un ensemble d'acteurs déterminés qui concourent à sa prise en charge

⁶ PTH : Prothèse Totale de Hanche

⁷ PTG : Prothèse Totale de genou

L'expérimentation se décompose en 2 phases, une première phase d'apprentissage dite « à blanc », qui dure 12 mois pour la colectomie pour cancer et 18 mois pour la PTH et PTG. Au cours de cette phase, les expérimentateurs mettront en œuvre leurs organisations et recueilleront les données nécessaires pour alimenter le modèle, où seront comparés, d'une part les dépenses réelles observées, et d'autre part, les paiements attendus correspondants aux forfaits cibles EDS. Et une deuxième phase de mise en œuvre, dite de financement rétrospectif, d'une durée de 3 ans. Cette phase sera appliquée de bonus si la dépense observée est inférieure au forfait cible, un intéressement pourra être versé à la condition qu'un niveau de qualité minimal soit atteint (mesuré à travers des indicateurs de qualités prédéfinis), ou de malus, qui pourront s'appliquer à la 3^{ème} année, si la dépense observée est supérieure au forfait cible et que la qualité n'a pas atteint le niveau minimal requis.

Les attendus de cette expérimentation se tiennent en 3 points. Le premier point, en termes d'amélioration du service rendu pour les usager en identifiant et maîtrisant les caractéristiques de leur état de santé qui peuvent constituer des facteurs de risque dans le cadre de la prise en charge, en anticipant les ruptures dans leur parcours, en leur proposant un accompagnement pluriprofessionnel et pluridisciplinaire cohérent tout au long de leur prise en charge, et ce dès la validation de l'indication de l'intervention chirurgicale et en prenant davantage en compte leurs retours relatifs à leur état de santé, leur qualité de vie, et leur expérience de prise en charge. Le deuxième point, en termes d'organisation et de pratiques pour les professionnels et les établissements ou services concourant à la prise en charge, en suscitant le développement de coopérations et d'organisations innovantes autour des prises en charge chirurgicales ciblées, prenant en compte les spécificités locales, en accélérant les actions déjà initiées autour d'une prise en charge globale du patient en amont, pendant et après l'intervention chirurgicale, et encourageant des relations plus fluides entre les professionnels mobilisés, en contribuant à renforcer le développement, l'harmonisation et le partage des bonnes pratiques entre les acteurs, et en participant à la maîtrise du risque de complications et à l'amélioration de l'état de santé du patient. Le troisième point, en terme d'efficience pour les dépenses de santé et de qualité des prises en charge, en contribuant à l'amélioration de la pertinence des soins, à l'amélioration de l'expérience patient par la préparation anticipée de son information et de sa prise en charge, la cohérence de l'ensemble de la prise en charge tout au long de l'épisode de soins selon son profil, son état de santé, ses facteurs de risque et ses conditions socio-économiques, à la diminution de la variabilité des pratiques, notamment dans l'orientation des patients en sortie d'hospitalisation, à la diminution des complications per et post interventions chirurgicales, et des hospitalisations potentiellement évitables, et à une meilleure allocation des dépenses de l'Assurance Maladie (AM) pour un épisode de soins.

4.2 Objectif et déroulement du stage

Dans ce contexte, l'objectif de mon stage est de vérifier si les hôpitaux volontaires à l'expérimentation, possèdent des spécificités, au niveau de leurs structures des réseaux hospitaliers, ainsi qu'au niveau du personnel hospitalier. Pour effectuer la première partie, j'ai utilisé des indicateurs, issus de la théorie des graphes, qui tente de résumer les réseaux des hôpitaux (degrés, liens, triades, centralité, etc...). Pour la seconde partie, j'ai utilisé des indicateurs territoriaux, qui résumé les données du personnel hospitalier, par département.

	S 1	S 2	S 3	S 4	S 5	S 6	S 7	S 8	S 9	S 10	S 11
<i>Prise en main du contexte</i>											
<i>Appropriation des données</i>											
<i>Analyses</i>											
<i>Présentation de nos analyses à l'équipe</i>											
<i>Rédaction du rapport de stage</i>											
<i>Préparation de l'oral de stage</i>											
<i>Réunion de département</i>											
<i>Réunion d'équipe</i>											

Tableau 1 : Planning du stage

Voici le déroulement du mon stage (*Tableau 1*). Les trois premières semaines m'ont permis de mieux comprendre le contexte, de m'informer sur le projet Episode De Soins, et de découvrir les premières bases de données. Par la suite, j'ai effectué mes analyses (descriptives, modélisation). J'ai pu participer à deux réunion avec l'équipe du projet, afin de présenter mes résultats, ainsi que deux réunions d'informations des personnels du département METIS, qui contribuaient à l'actualité de l'école, de présenter ces éventuels projets, et des problèmes rencontrés. Mon maître de stage était très présent durant mon stage, nous avons pu échanger plusieurs fois par semaines sur nos résultats, nos problématiques et ces solutions. Enfin, j'ai pu commencer à rédiger mon rapport de stage, dès la 5^{èmes} semaines de stage, ce qui m'a permis de prendre de l'avance sur sa rédaction.

Une partie de mes résultats, sera présenté lors du 58^{ème} colloques des Association de Science Régional en Langue Française, qui aura lieu les 29 et 30 juin, et le 1^{er} juillet 2022, à Rennes.

Session spéciale

Au-delà du paiement à l'acte.
Nouveaux modes de financement et organisation des soins de santé à l'échelle territoriale

L'expérimentation d'un paiement à l'épisode de soin à l'épreuve des territoires

Contributeurs: Anne-Laure Beaussier, Myriam Cluzel, Odessa Dariel, Anne Girault, Magali Robelet,
Nolwenn Le Meur, Hugo Maury, Alice Valiergue




Figure 3 : Capture d'écran de la trame de communication pour le 58ème colloques

Par la suite, dans une première partie, je vous présenterai une analyse sur la structure des réseaux hospitaliers et leurs impacts dans l'EDS., puis une analyse des indicateurs territoriaux des établissements. Je terminerai par présenter un bilan de nos recherches.

5 Analyse de la structure des réseaux hospitaliers et leurs impacts dans l'EDS

Je vais vous présenter mon analyse sur la structure des réseaux hospitaliers et leurs impacts dans l'EDS. Dans un premier temps, je vous présenterai le matériel que j'ai eu à disposition, puis la méthode que j'ai utilisée, afin de parvenir au résultat. Je terminerai par vous faire un bilan de cette recherche.

5.1 Matériels

Avant de faire mes analyses, j'ai effectué des analyses bibliographiques⁸, sur des documents que mon maître de stage m'a donné, afin de mieux comprendre les mots clés utilisés pour ce stage.

J'y ai également découvert plusieurs packages :

- SNA
- Network
- Igraph

Social Network Analyse (SNA), est un processus d'exploration de la structure sociale en utilisant la théorie des graphes. Il est utilisé pour mesurer et analyser les propriétés structurelles du réseau. Il permet de mesurer les relations et les flux entre les groupes, les organisations et les autres entités connectées.

Network peut être utilisée pour coder des structures relationnelles complexes composées d'un ensemble de vertex. Quelques fonctions simples de construction, d'interface et de visualisation sont fournis, ainsi qu'un ensemble d'opérateurs pour faciliter l'emploi par les utilisateurs finaux.

Igraph permet de fournir un ensemble de types de données et de fonctions pour la mise en œuvre d'algorithmes graphes, pour la manipulation rapide de grands graphes, avec des millions de sommets et de bords, et de permettre un prototypage rapide.

Les analyses statistiques ont été réalisées avec le langage de programmation R, un langage open source, avec l'IDE (environnement de développement informatique) Rstudio, version 4.0.5. J'ai utilisé plusieurs packages durant mon stage, dont les plus utilisés ont été dplyr, ggplot2, gtsummary et knitr⁹. J'ai exporté plusieurs documents de mes analyses, grâce à Rmarkdown.

⁸ Source 5 : Analyses Bibliographiques

⁹ Moteur de génération de rapport dynamiques avec R

Par soucis de confidentialité des données, les variables ont été générés grâce à l'Agence Technique de l'Information sur Hospitalisation (ATIH), puis calculé à l'aide du package *Igraph*, *SNA* et *Network* par mon maître de stage.

L'objectif de cette première analyse, est de comparer le profil des hôpitaux expérimentateurs et des hôpitaux non-expérimentateurs qui participent au projet EDS, grâce aux données de 2015 à 2019, au niveau des départements. Voici les variables indicatrices de réseaux, qui sont au nombres de 15, avec l'intervalle de leurs valeurs entre crochets :

- name : Identifiant de l'hôpital et son champ d'activité
- degree [0 , + Inf.] : Pour un hôpital, le nombre de liens vers d'autres établissements (échange d'au moins 1 patient = 1 lien), traduit par : *nombre de liens*
- strength_total [0 , + Inf.] : Nombre total de flux de patients, entrant et sortant, par hôpital (poids total des liens), traduit par : *flux patient*
- closeness_all [0 , 1] : L'autonomie/l'indépendance d'un hôpital, traduit par : *autonomie*
- centrality [0 , 1] : Notion d'importance d'un établissement dans un réseau, traduit par : *centralite*
- betweenness [0 , 1] : Interaction d'un hôpital avec d'autres établissements, traduit par : *interaction*
- prestige [0 , 1] : Attractivité d'un hôpital pour les patients d'un territoire, traduit par : *attractivite*
- pagerank [0 , 1] : Réputation d'un hôpital dans le territoire, traduit par : *reputation*
- triangles [binaire] : Participation d'un réseau de 3 hôpitaux, traduit par : *triades*
- finess : Identifiant géographique de l'hôpital
- experimentateur [binaire] : Hôpitaux participant, ou non, à l'expérimentation du projet Episode De Soins (0=non-expérimentateur, 1=expérimentateur)
- statuts : Statut juridique (publique, privé, à but non lucratif)
- reg17_label : Région de l'hôpital, parmi les 17 régions françaises
- annee : Année de collecte des données pour chaque hôpital
- Count : Nombre de fois qu'un hôpital est présent de 2015 à 2019

Définition des trois types d'établissements hospitaliers :

- MCO : Etablissement pratiquant les activités de médecine, chirurgie et obstétrique
- SSR : Etablissement pratiquant les activités de soins de suite et de réadaptation
- HAD : Etablissement pratiquant l'activité d'hospitalisation à domicile

Dans un premier temps, mon maître de stage m'a présenté 5 bases de données, qui concernaient, chacun, les indicateurs réseaux de l'années 2015 à 2019. Ensuite, j'ai effectué une jointure, en sélectionnant les établissements MCO. Voici la base de données que j'ai analysé :

name	strength	total_closeness	all_centrality	prestige	betweenness	pagerank	triangles	degree	finess	experimental	status	reg23_label	annee	Count
010000024_M	21	0.1927710843	0.1403508771	0.1927710843	0.2375	0.1139248213	1	10	10000024	0	MCO	Rhône-Alpes	2015	5
010000032_M	10	0.1797752808	0.1379310344	0.1797752808	0.1458333333	0.0853324866	0	8	10000032	0	MCO	Rhône-Alpes	2015	5
010000065_M	6	0.0663900414	0.0615384615	0.0663900414	0	0.0817965023	0	2	10000065	0	MCO	Rhône-Alpes	2015	4
010005239_M	10	0.1322314049	0.1032258064	0.1322314049	0.0958333333	0.0963934683	0	4	10005239	0	MCO	Rhône-Alpes	2015	5
010780195_M	12	0.1758241758	0.1481481481	0.0941176470	0.0958333333	0.0621887431	1	5	10780195	0	MCO	Rhône-Alpes	2015	5
010780203_M	6	0.1632653061	0.1367521367	0.1632653061	0.1270833333	0.0361833542	0	4	10780203	0	MCO	Rhône-Alpes	2015	5
020000089_M	4	0.2602739726	0.1958762886	0.1610169491	0.0730994152	0.0238405199	1	4	20000089	0	MCO	Picardie	2015	4
020000097_M	4	0.296875	0.2638888888	0.1809523809	0.0687134502	0.0248245631	1	4	20000097	0	MCO	Picardie	2015	1
020000105_M	5	0.2567567567	0.2289156626	0.1461538461	0.1023391812	0.0682689559	1	4	20000105	0	MCO	Picardie	2015	3
020000162_M	13	0.3220338983	0.2878787878	0.1809523809	0.4152046783	0.0701326367	1	9	20000162	0	MCO	Picardie	2015	5
020000246_M	5	0.2375	0.2209302325	0.1596638655	0.0994152046	0.0655548413	0	4	20000246	0	MCO	Picardie	2015	2
020000360_M	9	0.2794117647	0.2043010752	0.1759259259	0	0.0189887045	0	3	20000360	0	MCO	Picardie	2015	5
020000394_M	4	0.296875	0.2714285714	0.1809523809	0.1871345029	0.0375625763	0	4	20000394	0	MCO	Picardie	2015	5
020000519_M	13	0.3064516129	0.2261904761	0.1919191919	0.2719298245	0.0823305470	0	6	20000519	0	MCO	Picardie	2015	5
020000535_M	8	0.3064516129	0.2235294117	0.1809523809	0.1038011695	0.0411942955	1	6	20000535	0	MCO	Picardie	2015	5
020001061_M	9	0.1043956043	0.0935960591	0.0855855855	0	0.0643348803	0	2	20001061	0	MCO	Picardie	2015	5
020010047_M	11	0.2533333333	0.1958762886	0.1583333333	0	0.0460063660	1	5	20010047	0	MCO	Picardie	2015	5
030000061_M	11	0.3125	0.2419354838	0.2380952380	0.3976190476	0.0710306437	1	7	30000061	0	MCO	Auvergne	2015	5
030000079_M	7	0.3191489361	0.2586206896	0.25	0.5047619047	0.0398535278	1	7	30000079	0	MCO	Auvergne	2015	5
030000087_M	25	0.2830188679	0.2380952380	0.2173913043	0.3428571428	0.1324274076	1	9	30000087	0	MCO	Auvergne	2015	5
030002208_M	3	0.3	0.2112676056	0.2272727272	0.1285714285	0.0297883629	1	3	30002208	0	MCO	Auvergne	2015	4
030780548_M	3	0.2586206896	0.1456310679	0.1973684210	0.0761904761	0.0177705262	1	2	30780548	0	MCO	Auvergne	2015	4
030781116_M	35	0.3333333333	0.2678571428	0.2631578947	0.5	0.1309912450	1	8	30781116	0	MCO	Auvergne	2015	5
030785430_M	13	0.1666666666	0.1428571428	0.1388888888	0.0690476190	0.0612821421	0	5	30785430	0	MCO	Auvergne	2015	5
040000093_M	15	0.4210526315	0.2758620689	0.2857142857	0.4910714285	0.2072183234	1	9	40000093	0	MCO	Provence-Alp	2015	5
040000119_M	4	0.32	0.16	0.25	0	0.1004097513	1	3	40000119	0	MCO	Provence-Alp	2015	2
040000911_M	8	0.1904761904	0.1	0.1333333333	0	0.0655424308	0	2	40000911	0	MCO	Provence-Alp	2015	5
040780470_M	11	0.4444444444	0.3076923076	0.2424242424	0.4732142857	0.1170072376	0	6	40780470	0	MCO	Provence-Alp	2015	5

Tableau 2 : Extrait du fichier de données

5.2 Méthodes

Pour faire mes analyses, j'ai effectué des analyses descriptives des indicateurs, de 2015 à 2019, puis des hôpitaux expérimentateurs et non-expérimentateurs, afin d'observer d'éventuelles différences entre ces derniers. J'ai également réalisé une analyse univarié, avec une pvalue à hauteur de 5%, avec en Y, le fait qu'un établissement soit expérimentateur ou non-expérimentateur, au projet, afin d'observer l'impact de chaque variable. Puis j'ai utilisé le modèle ANOVA¹⁰ et le critère AIC¹¹, afin de construire un modèle logistique. Les résultats seront exprimés en Odds Ratios (OR), qui est le rapport des côtes d'exposition chez les établissements expérimentateurs et non-expérimentateurs. Enfin, j'ai utilisé une matrice de confusion, dans le but de vérifié la qualité du modèle de prédiction.

¹⁰ Analyse de la variance

¹¹ Mesure de qualité de modèle

5.3 Résultats

Dans un premier temps, j'ai comparé les hôpitaux expérimentateurs, aux hôpitaux non-expérimentateurs, année par année, afin de pouvoir constater, ou non, des évolutions dans le temps :

Variable	2015		2016		2017		2018		2019	
	expe, N = 30 [†]	n-expe, N = 792 [†]	expe, N = 30 [†]	n-expe, N = 792 [†]	expe, N = 30 [†]	n-expe, N = 792 [†]	expe, N = 30 [†]	n-expe, N = 792 [†]	expe, N = 30 [†]	n-expe, N = 792 [†]
flux patient	30 (15, 46)	12 (6, 21)	28 (17, 46)	12 (6, 21)	26 (16, 44)	11 (6, 20)	27 (14, 48)	11 (6, 20)	33 (19, 43)	11 (6, 20)
autonomie	0.36 (0.29, 0.41)	0.27 (0.19, 0.36)	0.37 (0.31, 0.42)	0.29 (0.21, 0.37)	0.35 (0.22, 0.44)	0.30 (0.21, 0.37)	0.37 (0.25, 0.43)	0.29 (0.21, 0.37)	0.36 (0.25, 0.42)	0.30 (0.22, 0.38)
centralité	0.15 (0.11, 0.22)	0.14 (0.11, 0.19)	0.17 (0.14, 0.19)	0.14 (0.11, 0.18)	0.17 (0.09, 0.23)	0.15 (0.11, 0.21)	0.14 (0.13, 0.22)	0.15 (0.12, 0.21)	0.13 (0.11, 0.22)	0.14 (0.10, 0.19)
attractivité	0.10 (0.07, 0.12)	0.10 (0.07, 0.15)	0.10 (0.07, 0.13)	0.10 (0.08, 0.14)	0.09 (0.07, 0.14)	0.11 (0.08, 0.15)	0.10 (0.08, 0.11)	0.11 (0.08, 0.15)	0.10 (0.07, 0.13)	0.12 (0.08, 0.17)
interaction	0.07 (0.04, 0.13)	0.06 (0.01, 0.16)	0.07 (0.03, 0.19)	0.06 (0.01, 0.16)	0.11 (0.05, 0.28)	0.06 (0.01, 0.15)	0.07 (0.03, 0.18)	0.06 (0.01, 0.16)	0.09 (0.03, 0.15)	0.06 (0.01, 0.15)
réputation	0.04 (0.01, 0.06)	0.03 (0.02, 0.07)	0.03 (0.02, 0.06)	0.03 (0.02, 0.07)	0.04 (0.02, 0.08)	0.03 (0.01, 0.07)	0.03 (0.02, 0.06)	0.03 (0.02, 0.07)	0.04 (0.02, 0.06)	0.03 (0.01, 0.07)
triades	27 (90%)	586 (74%)	28 (93%)	577 (73%)	28 (93%)	580 (73%)	28 (93%)	602 (76%)	29 (97%)	575 (73%)
nombre de liens	13 (8, 20)	6 (4, 10)	14 (9, 20)	7 (4, 10)	16 (11, 22)	6 (4, 10)	14 (9, 18)	6 (4, 10)	16 (9, 21)	6 (4, 10)

[†] Median (IQR); n (%)

Légende :

n-expe : correspond aux nombres d'établissements non-expérimentateurs par année

expe : correspond aux nombres d'établissements expérimentateurs par année

□ : Souligne l'absence de contraste

Tableau 3 : Comparaison, année par année, des indicateurs réseaux des hôpitaux expérimentateurs et non-expérimentateurs

Au vu de cette première approche, j'ai pu conclure que dès 2015, il semble y avoir des différences entre ces deux parties, comme le nombre liens ou le flux de patient. De plus, ces différences restent stables au fur et à mesure des années, comme vous pouvez le voir dans les deux encadrés en rouge. Les hôpitaux expérimentateurs ne semblent pas avoir changé de positionnement au sein de leurs réseaux, tout comme les hôpitaux non-expérimentateurs.

N'ayant pas observé d'impact temporel, j'ai alors comparé ces indicateurs, sans prendre en compte la notion des années. J'ai obtenu le tableau suivant :

Variable	Comparaison		Modele univarie	
	expe, N = 150 ¹	n-expe, N = 3,960 ¹	OR ²	p-value
flux patient	29.50 (15.25, 45.75)	11.00 (6.00, 21.00)	1.04	<0.001
autonomie	0.36 (0.27, 0.42)	0.29 (0.21, 0.37)	42.1	<0.001
centralite	0.15 (0.12, 0.22)	0.15 (0.11, 0.20)	6.45	0.042
attractivite	0.10 (0.07, 0.13)	0.11 (0.08, 0.15)	0.25	0.3
interaction	0.08 (0.03, 0.19)	0.06 (0.01, 0.15)	2.98	0.046
reputation	0.04 (0.02, 0.06)	0.03 (0.02, 0.07)	2.65	0.5
triades	140 (93%)	2,920 (74%)	4.99	<0.001
nombre de liens	14.00 (9.00, 20.00)	6.00 (4.00, 10.00)	1.13	<0.001
¹ Median (IQR); n (%)				
² OR = Odds Ratio				

Légende :

no : correspond aux nombres d'établissements non-expérimentateurs, de 2015 à 2019

yes : correspond aux nombres d'établissements expérimentateurs, de 2015 à 2019

 : Souligne les différences observées

Tableau 4 : Comparaison des indicateurs réseaux, sur la période 2015 – 2019, des hôpitaux expérimentateurs et non-expérimentateurs

La partie *Comparaison* confronte les indicateurs de structure de réseaux, du nombre total d'hôpitaux expérimentateurs, au nombre total d'hôpitaux non-expérimentateurs, toutes années confondues. Concernant la partie *Modele*, *OR* correspond à l'Odds Ratio¹² de chaque variable, en utilisant un modèle univarié¹³, enfin, *pvalue* mesure la significativité de chaque variable dans le modèle.

Lors de cette deuxième approche, j'ai pu constater des différences au niveau de la structure de réseaux entre les hôpitaux expérimentateurs et non-expérimentateurs. En effet, les hôpitaux expérimentateurs ont tendances à être mieux encrés et posséder plus de liens et d'échanges de patients avec d'autres hôpitaux, que les hôpitaux non-

¹² Impact d'une variable dans un modèle de régression

¹³ Modèle de régression utilisant 1 seul paramètre

expérimentateurs. J'ai également pu constater que les hôpitaux expérimentateurs ont tendance à être plus autonome.

Par la suite, j'ai utilisé une matrice de corrélation, afin d'observer la présence d'éventuelle colinéarité entre variables.

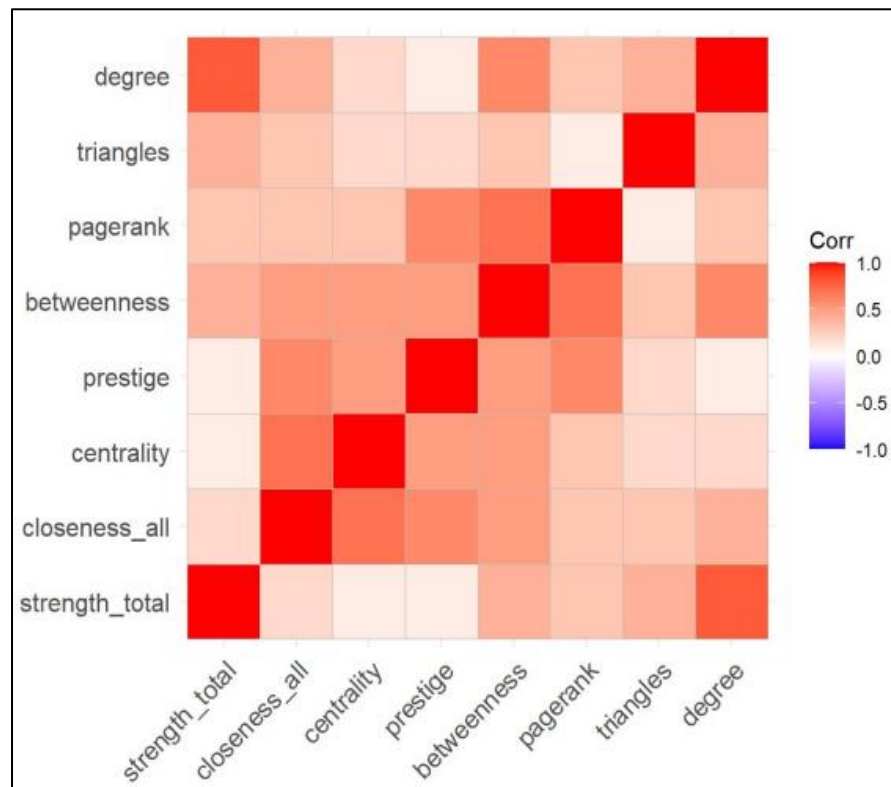


Figure 4 : Matrice de corrélation des variables

J'ai pu observer une forte association entre la variable *degree* et *strength_total*, ce qui n'est pas une coïncidence puisque les deux variables sont intimement liées (*degree* mesure le nombre d'échange de patient par hôpital, et *strength_total* mesure le nombre de patients, entrant et sortant, par hôpital). Avec mon maître de stage, nous avons décidé de retirer la variable *degree*.

Après avoir comparé la structure des réseaux d'hôpitaux expérimentateurs et non-expérimentateurs, j'ai mis en place une régression logistique, afin de pouvoir prédire si un hôpital est expérimentateur ou non. Dans un premier temps, j'ai découpé l'échantillon en deux parties : un échantillon d'apprentissage et un échantillon test. L'échantillon d'apprentissage sera utilisé pour le modèle logistique, puis j'ai utilisé l'échantillon test, afin de tester la précision du modèle. J'ai commencé par créer un

modèle complet ¹⁴ (en ayant retiré la variable *degree*), puis j'ai vérifié la significativité des variables, grâce au modèle ANOVA ¹⁵ :

```
## Analysis of Deviance Table (Type II tests)
##
## Response: experimentateur
##          LR Chisq Df Pr(>Chisq)
## closeness_all    28.159  1 1.118e-07 ***
## centrality        0.208  1  0.648323
## prestige         11.673  1  0.000634 ***
## betweenness       6.600  1  0.010199 *
## pagerank          0.133  1  0.715849
## triangles         7.333  1  0.006771 **
## strength_total   76.779  1 < 2.2e-16 ***
## factor(annee)     0.164  4  0.996819
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 5: ANOVA du modèle logistique complet

Au vu des pvalues, j'ai constaté la présence de variables non significatives. J'ai alors éliminé, de manière automatique, les variables non pertinentes selon le critère AIC (Akaike Information Criterion) ¹⁶, et j'ai obtenu le résultat suivant :

```
## Call:
## glm(formula = experimentateur ~ closeness_all + prestige + betweenness +
##      triangles + strength_total, family = binomial(link = "logit"),
##      data = dataset2)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -1.6638  -0.2800  -0.2067  -0.1489   3.4277
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -5.61221    0.39819  -14.094 < 2e-16 ***
## closeness_all    5.63272    0.97087   5.802 6.56e-09 ***
## prestige       -5.90903    1.74283  -3.390 0.000698 ***
## betweenness    -2.35294    0.83295  -2.825 0.004731 **
## triangles        0.84655    0.34606   2.446 0.014434 *
## strength_total  0.03466    0.00365   9.498 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1287.6  on 4109  degrees of freedom
## Residual deviance: 1112.5  on 4104  degrees of freedom
## AIC: 1124.5
```

Figure 6: Modèle logistique final

¹⁴ Modèle de régression comprenant toute les variables

¹⁵ Analyse de la variance

¹⁶ Mesure de qualité de modèle

Afin de tester notre modèle logistique, j'ai utilisé une matrice de confusion, qui m'a permis de constater sa qualité de prédiction. Une matrice de confusion est une matrice qui mesure la qualité d'un système de classification. Chaque ligne correspond à une classe réelle, chaque colonne correspond à une classe estimée. Elle ne compare donc que les données réelles pour une variable cible, à celles prédites par un modèle.

J'ai appliqué notre modèle de prédiction sur l'échantillon test, et la matrice de confusion les analysera, pour pouvoir constater si le modèle est performant, ou non.

Les indicateurs principaux :

- Accuracy : Taux de bon classement du modèle
- Sensitivity : Sa sensibilité du modèle, c'est-à-dire le taux de bien classé parmi les expérimentateurs qui ont participé à l'étude
- Specificity : Spécificité du modèle, c'est à dire le taux de bien classé parmi les expérimentateurs qui n'ont pas participé à l'étude
- Pos Pred Value : Précision du modèle, c'est à dire le taux de bien classés parmi les expérimentateurs qui ont participé à l'étude classés comme expérimentateurs qui ont participé à l'étude
- Neg Pred Value : Taux de bien classés parmi les expérimentateurs classés comme non participant
- Prevalence : Taux d'établissements expérimentateurs qui ont réellement participé à l'étude

Tout d'abord, j'ai cherché le seuil de probabilité optimal, permettant de considérer qu'un établissement soit un expérimentateur. J'ai trouvé un compromis entre le taux de vrais positifs, c'est à dire la proportion de vrais positifs parmi les classés positifs ou sensibilité, et le taux de vrais négatifs, c'est à dire la proportion de vrais négatifs parmi les classés négatifs ou spécificité. J'ai obtenu le graphique suivant :

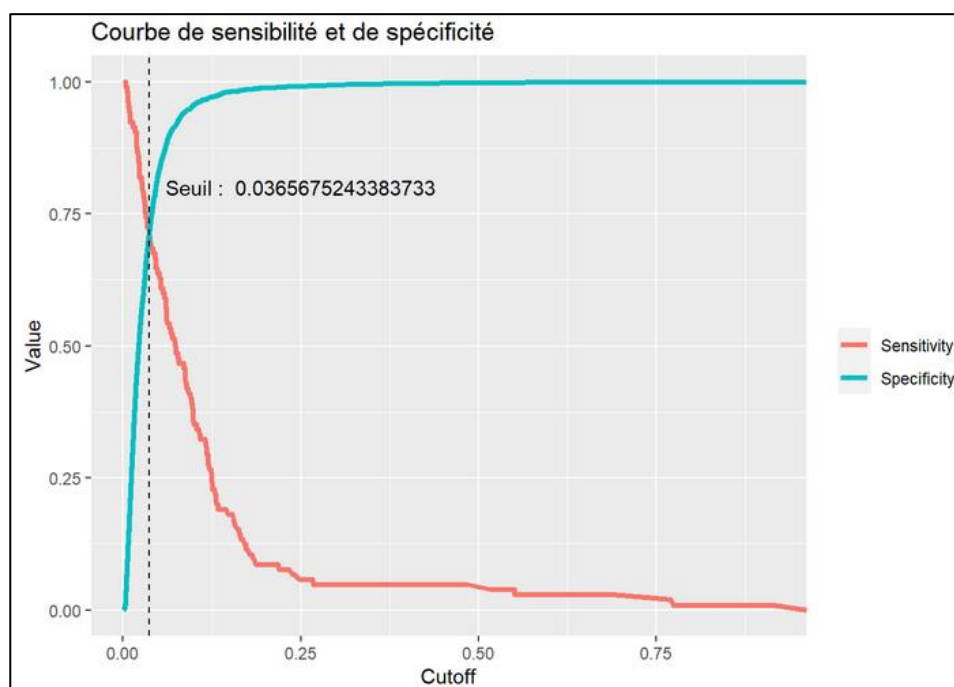


Figure 7 : Courbe de sensibilité et de spécificité

J'ai donc conservé le seuil à 3,7%, afin d'optimiser un maximum la sensibilité et la spécificité, puis j'ai effectué la matrice de confusion suivante :

Confusion Matrix and Statistics		
Reference		
Prediction	no	yes
no	832	11
yes	356	34
Accuracy : 0.7024		
95% CI : (0.676, 0.7278)		
No Information Rate : 0.9635		
P-Value [Acc > NIR] : 1		
Kappa : 0.0972		
McNemar's Test P-Value : <2e-16		
Sensitivity : 0.75556		
Specificity : 0.70034		
Pos Pred Value : 0.08718		
Neg Pred Value : 0.98695		
Prevalence : 0.03650		
Detection Rate : 0.02758		
Detection Prevalence : 0.31630		
Balanced Accuracy : 0.72795		
'Positive' Class : yes		

Tableau 5 : Matrice de confusion

Grâce à l'indicateur « Accuracy », notre modèle logistique prédit environ 70% de notre échantillon test. Nous avons pu constater qu'un nombre important d'hôpitaux non-expérimentateurs sont prédit en tant qu'expérimentateur (356), et l'inverse est également observé. Nous avons supposé certains des hôpitaux non-expérimentateurs possédaient des caractéristiques pour être expérimentateur, et à l'inverse, certains des hôpitaux expérimentateurs possédaient des caractéristiques pour ne pas l'être.

Enfin, j'ai réalisé un graphe, permettant de représenté les réseaux de transfère des patients EDS entre structures de soins, à la suite d'une prise en charge orthopédique, dans un établissement expérimentateur :

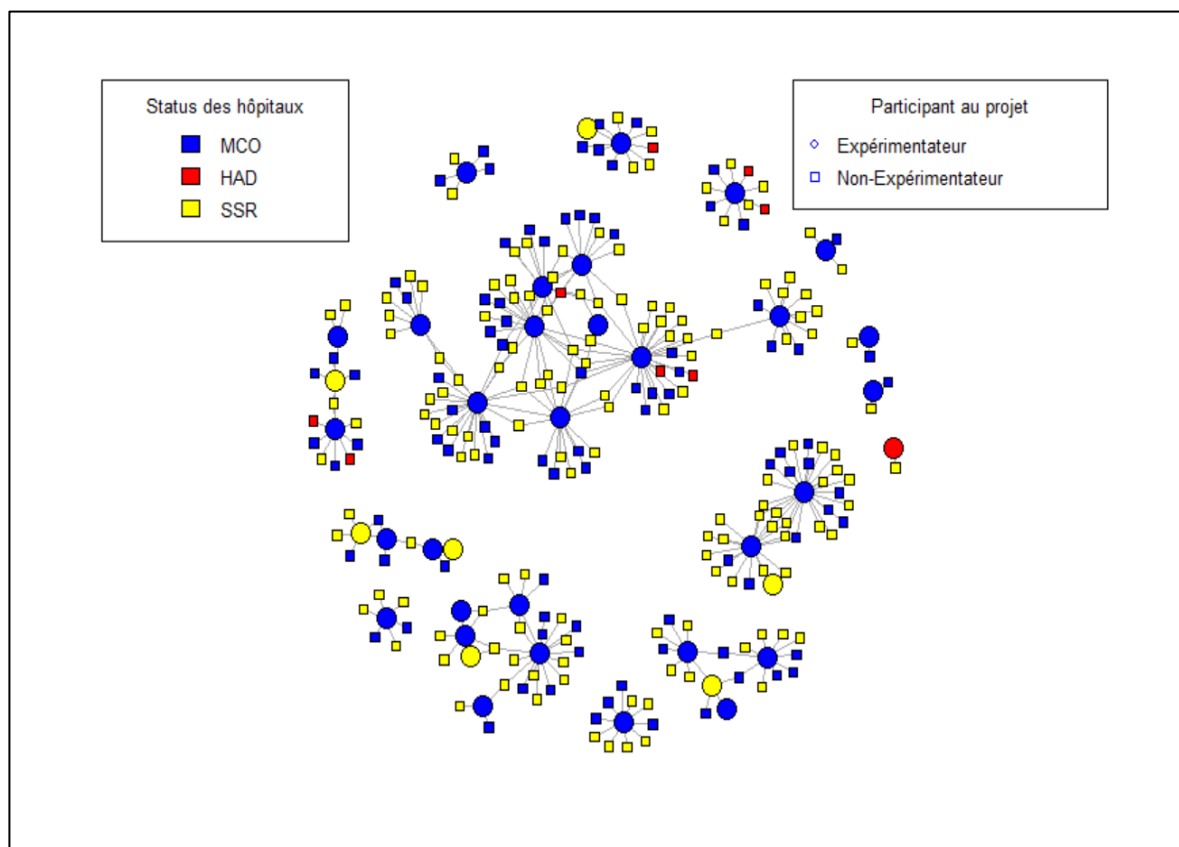


Figure 8: Graphe représentant les réseaux de transfère des patients, venant des hôpitaux expérimentateurs

Nous avons pu observer que la majorité des hôpitaux expérimentateurs, possédaient un bon réseau de soins. Les établissements expérimentateurs MCO, ont tendance à transférer la majorité de leurs patients vers des établissements SSR. Ces derniers sont très représentés dans les réseaux, tout au contraire des HAD, qui ne semblent pas être très présent.

5.4 Conclusion

Pour conclure sur cette partie, nous avons pu constater que les hôpitaux expérimentateurs ont tendances à posséder un plus vaste réseau de soin, et être mieux intégrer, que les hôpitaux non-expérimentateurs, ce qui peut créer un problème, puisque ces derniers semblent moins préparés pour le projet. D'après le modèle de régression logistique, les variables concernant le nombre de patient, l'autonomie et les triades, semblent expliquer si un établissement peut être expérimentateur ou non.

Après avoir analysé la structure des réseaux hospitaliers et leurs impacts dans l'EDS, nous nous sommes penché sur les indicateurs territoriaux des établissements expérimentateurs.

6 Analyse des indicateurs territoriaux des établissements

Dans cette deuxième partie, je vais vous présenter l'analyse sur les indicateurs territoriaux des établissements expérimentateurs. Dans un premier temps, je vous présenterai le matériel que j'ai eu à disposition, puis la méthode que j'ai utilisée, afin de parvenir au résultat. Je terminerai par vous faire un bilan de cette recherche.

6.1 Matériels

A l'aide d'une simulation des forfaits, du projet Episode De Soins, l'équipe a pu déterminer les établissements, dont le forfait serait avantageux, et les établissements, dont le forfait serait désavantageux, en fonction de sa pratique (prothèse totale de hanche, prothèse totale de genou) et au niveau des pratiques globales.

J'ai récupéré plusieurs bases de données, que mon maître de stage a généré à partir d'atlas s@nté, afin de les triées, et de pouvoir sélectionner les données pertinentes pour cette analyse. J'ai pu créer plusieurs bases de données, à partir des données trier, afin de comparer ces indicateurs en fonction des caractéristiques de chaque établissement (expérimentateur/non-expérimentateur, établissement estimé gagnant/perdant, patient complexe/non-complexe, prothèse totale de hanche/prothèse totale de genou), dont voici les variables principales que j'ai pu utiliser, avec l'intervalle des valeurs entre crochets :

- gagnant [binaire] : Indique si la simulation des forfaits est avantageuse, ou non, pour chaque établissement (0 = perdant, 1 = gagnant)
- pth [binaire] : Indique si la simulation des forfaits est avantageuse, ou non, pour les établissements pratiquants les prothèses totales pour hanches (0 = perdant, 1 = gagnant)
- ptg [binaire] : Indique si la simulation des forfaits est avantageuse, ou non, pour les établissements pratiquants les prothèses totales pour genoux (0 = perdant, 1 = gagnant)
- tamax_chir [0 , + Inf.] : Temps d'accès maximum (en minute), à une structure sanitaire ou médico-sociale la plus proche en 2021, en service Chirurgie, par département
- tamiax_ssr [0 , + Inf.] : Temps d'accès maximum (en minute), à une structure sanitaire ou médico-sociale la plus proche en 2021, en service de soins de suite, par département

- tamax_longstay [0 , + Inf.] : Temps d'accès maximum(en minute), à une structure sanitaire ou médico-sociale la plus proche en 2021, en service de longs séjours, par département
- tamax_ssiad [0 , + Inf.] : Temps d'accès maximum (en minute), à une structure sanitaire ou médico-sociale la plus proche en 2021, en service de Soins Infirmiers A Domicile, par département
- tamax_rhumato [0 , + Inf.] : Temps d'accès maximum (en minute), à un professionnel libéral en 2021, en rhumatologie, par département
- nb_csip [0 , + Inf.] : Nombre de Centre de Soins Infirmiers et Polyvalent, par département, en 2021
- nb_ssiad [0 , + Inf.] : Nombre de Service de Soins Infirmiers A Domicile, par département, en 2021
- density_nurse [0 , + Inf.] : Nombre d'infirmier pour 10 000 habitants, en 2019, par département
- density_kine [0 , + Inf.] : Nombre de kinésithérapeute pour 10 000 habitants, en 2019, par département
- Taux_cpx0 [0 , 100] : Patient ne possédant pas de complexité, ayant subi une opération orthopédique dans un établissement MCO, puis transféré dans un établissement SSR
- Taux_cpx1 [0 , 100] : Patient possédant une complexité, ayant subi une opération orthopédique dans un établissement MCO, puis transféré dans un établissement SSR

Voici un extrait de la base de données que j'ai analysé :

finess	departement	gagnant	pth	ptg	tamax_chir	tamax_ssr	tamax_long	tamax_ssiad	tamax_rhum	nb_csip	nb_ssiad	density_nurs	density_kine	Taux_cpx0	Taux_cpx1
250000270	25	1	1	1	65	52	65	29	63	2	3	11.7687074	6.19319727	28.2	77.78
310780150	31	0	0	0	52	36	48	24	52	126	71	23.7330049	14.3379310	40.8	70.83
310783048	31	0	0	0	45	45	45	34	45	6	10	21.9544117	13.15	38.87	51.85
330000340	33	1	1	0	47	43	33	21	46	4	3	17.8260869	13.4304347	38.46	76.92
330781360	33	1	1	1	44	34	46	26	44	56	12	16.13	9.3475	21.46	50
330782582	33	1	1	1	40	40	47	19	31	4	4	18.7266666	14.12	14.73	100
340780634	34	0	0	0	30	22	32	25	15	37	20	18	27.55	56.74	59.09
350000121	35	1	1	1	38	23	31	20	38	35	11	10.6690476	7.77619047	16.79	51.06
400780342	40	0	0	0	30	28	30	34	28	0	1	19.3181818	18.9727272	37.93	NA
420010050	42	0	0	0	42	25	46	18	49	37	37	16.8585365	11.1341463	47.34	73.86
570001057	57	1	1	1	43	35	36	35	61	0	1	13.4969696	6.0454545	17.3	29.37
590000618	59	1	1	1	25	15	28	14	28	5	8	15.1571428	10.7928571	38.21	64.94
590780383	59	0	0	0	20	20	19	10	13	85	18	12.575	13.9083333	40.47	63.83
590781951	59	1	1	1	19	18	19	12	19	87	24	12.8666666	14.8388888	24.75	45.45
640780490	64	0	1	0	12	17	17	17	12	3	1	30.98	31.24	39.66	33.33
690780259	69	0	1	0	32	32	32	31	49	12	8	13.165	9.52	44.48	75
690780648	69	1	1	1	40	32	36	21	60	9	8	12.6923076	7.26410256	17.31	33.33
690780655	69	0	1	0	8	11	19	8	8	7	3	13.65	8.75	46.67	83.33
710978347	71	1	0	1	35	20	52	20	35	2	2	17.1785714	5.06428571	44.29	42.86
750000523	75	1	1	1	61	22	45	42	51	63	20	13.2320754	8.16226415	25.71	48.15
750100042	75	0	1	0	18	12	19	14	14	140	25	6.87058823	11.1235294	45.67	50
750150104	75	0	0	0	27	27	35	14	27	72	24	7.82222222	7.21666666	28.11	36.07
750150237	75	1	1	0	45	40	54	25	47	130	38	9.940625	6.66145833	26.77	60.87
750300493	75	0	1	0	95	75	95	84	95	105	30	28.3592233	13.2320388	22.14	38.71
800006124	80	1	1	1	50	39	38	27	54	12	10	12.5225641	6.33076923	19.18	75
830100574	83	0	0	0	56	40	56	28	45	12	18	25.2411764	14.3176470	61.03	50
900000035	90	1	0	1	34	25	31	18	34	3	4	9.62972972	4.09729729	36.52	61.54
920100013	92	0	1	0	24	24	38	19	24	65	17	6.946	6.62	30.47	36.36
940120017	94	0	1	0	14	14	18	8	0	12	5	6.425	12.975	43.79	60
950032714	95	1	1	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	2.17	73.68
950007982	95	1	1	1	23	16	24	20	18	11	5	8.21515151	6.84242424	28.5	0

Tableau 6 : Extrait du fichier de données

L'objectif de cette seconde partie, est, de comparer les profils « gagnants » et « perdants », d'une rémunération au forfait, parmi les 30 établissements expérimentateurs.

6.2 Méthodes

J'ai effectué des analyses descriptives des données territoriales, concernant, dans une première partie, les établissements recensés expérimentateurs, en 2019, puis sur l'ensemble des établissements expérimentateurs et non-expérimentateurs, de 2015 à 2019. J'ai recensé 30 établissements expérimentateurs, dont 15 sont estimés gagnants et 14 sont estimés perdant, au niveau global, 12 sont estimés gagnants et 17 sont estimés perdants, au niveau du type de soin concernant la prothèse totale de genou (PTG), et 20 sont estimés gagnants et 10 sont estimés perdants, au niveau du type de soin concernant la prothèse totale de hanche (PTH). Je n'ai pas observé de données, concernant un établissement, au le niveau global et du type de soin sur la prothèse de soin de genou. Je n'ai donc pas sélectionné cet établissement dans ces deux analyses.

Afin de déterminer si un établissement est gagnant ou perdant, au niveau global, j'ai additionné le coût réel et le coût théorique du forfait, au niveau des deux types de soin (prothèse totale de genou, puis de hanche), pour chaque établissement. Si le résultat était négatif, l'établissement est donc perdant, et au contraire, si le résultat est positif, l'établissement est gagnant.

6.3 Résultats

Dans un premier temps, j'ai effectué mes analyses sur les établissements gagnants et perdants, au niveau global, puis dans un second temps, au niveau des types de pratique (prothèse totale de genou et prothèse totale de hanche).

Pour commencer, j'ai représenté la densité de kinésithérapeutes, en fonction des établissements gagnant et perdant. J'ai obtenu le graphique suivant :

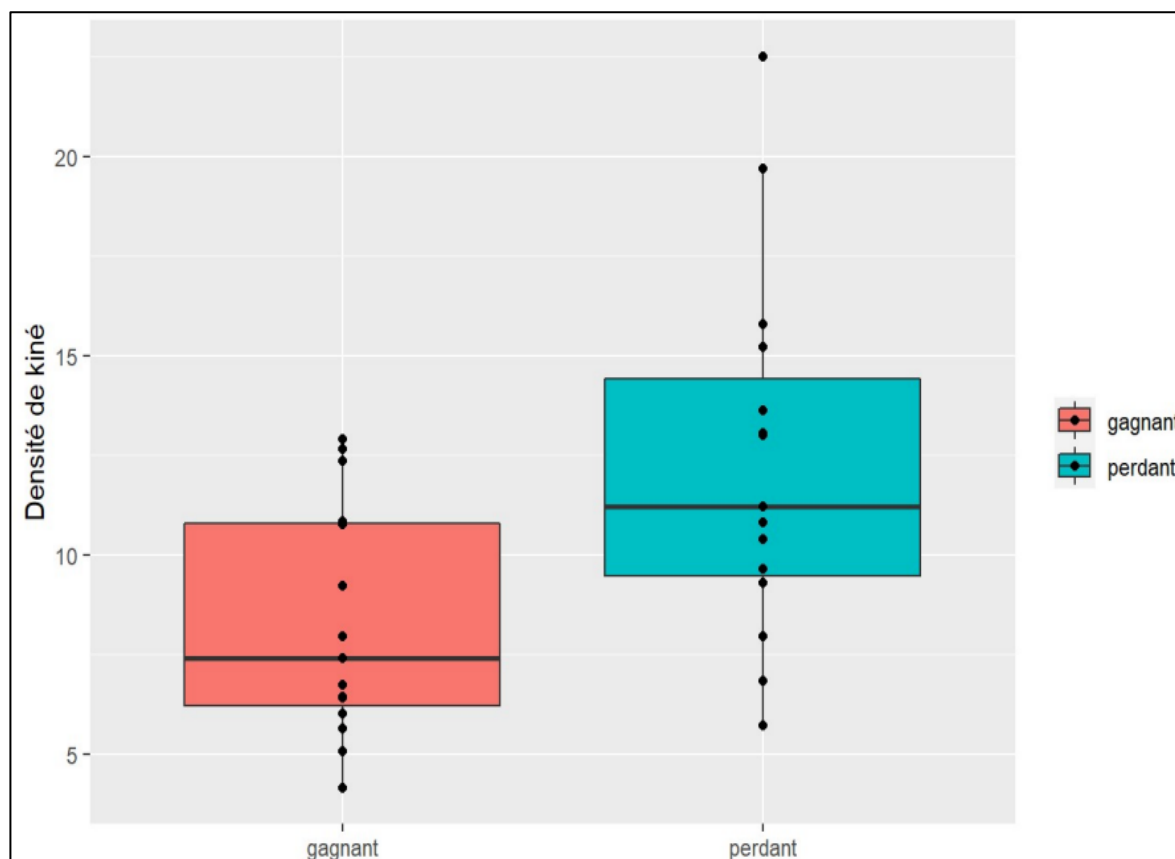


Figure 9 : Distribution de la densité de kinésithérapeutes, pour 10 000 habitants, à partir des simulations des établissements gagnants/perdants, en 2019

Au vu de ce graphique, 50% des territoires de la patientèle des hôpitaux perdants, ont tendances à posséder, approximativement, deux fois plus de kinésithérapeutes, que la plupart des territoires des hôpitaux gagnants, pour l'année 2019.

Par la suite, j'ai comparé les indicateurs territoriaux des territoires possédants un établissement gagnant et perdant, afin de constater la présence, ou non, de différence entre ces derniers. J'ai obtenu le tableau suivant :

Variable	Comparison		univariate model	
	losing, N = 14 ¹	winning, N = 15 ¹	OR ²	p-value
tamax_chir	28.50 (18.50, 44.25)	40.00 (34.50, 46.00)	1.02	0.333
tamiaux_ssr	24.50 (17.75, 35.00)	32.00 (21.00, 39.50)	1.01	0.741
tamax_longstay	33.50 (19.00, 45.75)	36.00 (31.00, 46.50)	1.01	0.755
tamax_ssiad	18.50 (14.00, 27.25)	21.00 (19.50, 26.50)	1.00	0.909
tamax_rhumato	25.50 (13.25, 48.00)	44.00 (32.50, 52.50)	1.03	0.203
nb_csip	37.00 (12.00, 81.75)	9.00 (3.50, 45.50)	0.99	0.165
nb_ssiad	18.00 (8.50, 24.75)	8.00 (3.50, 11.50)	0.94	0.090
density_nurse	15.25 (9.01, 23.29)	12.87 (11.22, 15.64)	0.91	0.173
density_kine	13.06 (9.92, 14.22)	7.26 (6.26, 10.07)	0.74	0.027
Taux_cpx0	42.30 (39.07, 46.42)	25.71 (18.24, 32.51)	0.86	0.004
Taux_cpx1	55.47 (41.53, 69.08)	51.06 (44.16, 69.97)	1.00	0.851
¹ Median (IQR)				
² OR = Odds Ratio				
<p style="text-align: center;"><u>Légende :</u></p> <p style="text-align: center;"><i>losing</i> = établissement estimé perdant <i>winning</i> = établissement estimé gagnant : Souligne les différences observées</p>				

Tableau 7 : Comparaison des indicateurs territoriaux, à partir de la simulation des établissements de soins estimés gagnants et perdants

Comme encadré en rouge (tableau 7), il semble avoir plus de kinésithérapeutes, dans les territoires des établissements perdants, que gagnants. De même pour les Centre de Soins Infirmier Polyvalent ¹⁷et les Services de Soins Infirmiers A Domicile¹⁸. Concernant le taux d'adressage, je n'ai pas observé de différence au niveau des patients ayant de lourdes comorbidités. Cependant, les territoires d'attractivités des établissements perdants ont tendances à transférer davantage les patients ne possédant pas de complexité, que les territoires des établissements gagnants. Les Odds Ratio, de

¹⁷ CSIP

¹⁸ SSIAD

l'analyse univarié, confirme les résultats, puisque la plupart sont en inférieur au seuil 1, ce qui signifie que lorsque l'on passe d'un territoire possédant un établissement estimé perdant, à un territoire possédant un établissement estimé gagnant, la tendance sera d'observer des baisses au niveau des structures et du personnel hospitalier. Enfin, il ne semble pas y avoir de différence, au niveau du temps d'accès maximum, à une structure de soin. Au vu de ces résultats, nous avons pu supposer, que les territoires des établissements perdants, possèdent plus de patients, et donc plus de personnels et de structures hospitalières, que les territoires des établissements gagnants. De plus, nous pouvons également faire l'hypothèse que les territoires possédant un établissement perdant, seraient capables d'accueillir plus de patients, et posséder plus de personnels, et donc d'augmenter la probabilité d'observer un patient ayant une complexité, ce qui pourrait confirmer notre première hypothèse. La complexité des patients semble être un des indicateurs, permettant de définir si le forfait serait avantageux, ou non, pour chaque établissement.

Par la suite, j'ai appliqué la même méthode, en fonction du type de soin, afin de comparer si un des types seraient avantageux, ou non. J'ai obtenu les résultats suivant :

Au niveau du type de soin concernant la prothèse totale de hanche :

Variable	Comparison		Univariate model	
	losing, N = 10 ¹	winning, N = 20 ¹	OR ²	p-value
tamax_chir	34.500 (30.000, 44.250)	39.000 (22.000, 45.500)	1.00	0.994
tamiaux_ssr	26.000 (22.750, 34.000)	28.000 (16.750, 39.250)	1.00	0.864
tamax_longstay	40.000 (31.250, 47.500)	34.500 (22.750, 45.250)	0.99	0.706
tamax_ssiad	22.000 (18.000, 27.250)	20.500 (16.250, 27.500)	1.01	0.698
tamax_rhumato	34.500 (27.250, 45.000)	41.000 (18.750, 51.750)	1.01	0.626
nb_csip	24.500 (3.750, 63.250)	12.000 (4.750, 63.500)	1.00	0.995
nb_ssiad	18.000 (5.500, 23.000)	8.000 (3.750, 17.750)	0.96	0.167
density_nurse	17.589 (13.646, 21.295)	13.016 (10.487, 15.400)	0.92	0.189
density_kine	13.529 (8.196, 14.333)	9.049 (6.797, 13.039)	0.94	0.318
¹ Median (IQR)				
² OR = Odds Ratio				

Légende :

losing = établissement estimé perdant

winning = établissement estimé gagnant

: Souligne les différences observées

Tableau 8 : Comparaison des indicateurs territoriaux, concernant le type de soin sur la prothèse totale de hanche, à partir des établissements estimés gagnants et perdants

Concernant la prothèse totale de hanche, les territoires des établissements estimés gagnants à ce niveau-là, ont tendance à posséder moins de structures hospitalières (CSIP, SSIAD), que les territoires des établissements estimés perdants, même si statistiquement cela n'est pas significatif. De plus, les temps d'accès maximum ne semblent pas les différencier. Concernant l'analyse univarié, aucune des variables possèdent une pvalue inférieur à 5%, ce qui ne signifie qu'aucune de ces variables expliquent, significativement, le fait qu'un territoire possède un établissement gagnant ou perdant, concernant la prothèse totale de hanche.

Au niveau du type de soin concernant la prothèse totale de genou :

Variable	Comparison		Univariate model	
	losing, N = 17 ¹	winning, N = 12 ¹	OR ²	p-value
tamax_chir	29.000 (24.000, 41.000)	33.000 (25.500, 43.000)	1.01	0.690
tamiaux_ssr	26.000 (23.000, 33.000)	24.500 (20.250, 34.250)	1.01	0.815
tamax_longstay	31.000 (27.000, 46.000)	29.500 (23.750, 38.500)	0.97	0.356
tamax_ssiad	23.000 (20.000, 28.000)	23.000 (18.000, 28.500)	1.01	0.913
tamax_rhumato	34.000 (22.000, 46.000)	34.500 (22.500, 45.750)	1.00	0.948
nb_csip	41.000 (11.000, 102.000)	11.000 (4.750, 44.250)	0.98	0.091
nb_ssiad	23.000 (8.000, 34.000)	10.000 (6.750, 17.750)	0.96	0.136
density_nurse	15.868 (8.961, 21.437)	13.293 (10.091, 17.865)	0.95	0.387
density_kine	12.187 (7.968, 13.618)	7.743 (6.404, 11.732)	0.84	0.113
¹ Median (IQR)				
² OR = Odds Ratio				

Légende :

losing = établissement estimé perdant

winning = établissement estimé gagnant

: Souligne les différences observées

Tableau 9 : Comparaison des indicateurs territoriaux, concernant le type de soin sur la prothèse totale de genou, à partir des établissements estimés gagnants et perdants

Concernant la prothèse totale de genou, j'ai pu retrouver les mêmes résultats que ceux obtenus précédemment. Les territoires des établissements estimés perdant semblent posséder plus de structures de soin. L'analyse univariée informe également qu'aucunes variables n'est significatives, et peut expliquer le fait qu'un territoire possède un établissement gagnant ou perdant.

Par la suite, j'ai effectué un modèle logistique, dans le but de constater, si un modèle multivarié, pourrait prédire qu'un territoire possède un établissement gagnant ou perdant, au niveau global. Dans un premier temps, j'ai sélectionné les 5 variables qui semblait expliquer le mieux ce phénomène (*nb_ssiad*, *nb_csip*, *density_nurse*, *density_kine* et *Taux_cpx0*), en ayant un pvalue inférieur à 20%. Ensuite, j'ai utilisé le critère AIC, afin d'avoir un modèle de prédiction optimal. Voici ce que j'ai obtenu :

Variable	Comparison		Univariate model		Multivariate model	
	losing, N = 14 ¹	winning, N = 15 ¹	OR ²	p-value	OR ²	p-value
nb_csip	37.00 (12.00, 81.75)	9.00 (3.50, 45.50)	0.99	0.165		
nb_ssiad	18.00 (8.50, 24.75)	8.00 (3.50, 11.50)	0.94	0.090	0.85	0.095
density_nurse	15.25 (9.01, 23.29)	12.87 (11.22, 15.64)	0.91	0.173		
density_kine	13.06 (9.92, 14.22)	7.26 (6.26, 10.07)	0.74	0.027	0.82	0.186
Taux_cpx0	42.30 (39.07, 46.42)	25.71 (18.24, 32.51)	0.86	0.004	0.77	0.049
¹ Median (IQR)						
² OR = Odds Ratio						

Légende :

losing = établissement estimé perdant

winning = établissement estimé gagnant

□ : Souligne l'absence de significativité

Tableau 10 : Odds Ratio de la régression logistique multivariée, à partir de la simulation des établissements de soins estimés gagnants et perdants, au niveau global

Au vu des pvalues, pour le modèle multivarié, ce modèle n'explique pas l'estimation des établissements gagnants et perdants. La pvalue de *density_kine* peut supposer une absence de comorbidités lourdes.

Ce que l'on peut retenir, à propos du modèle univarié, et multivarié, seul le nombre de kinésithérapeutes pour 10 000 habitants, et le taux d'adressage de patient n'ayant pas de lourdes comorbidités, venant d'établissements MCO, vers des établissements SSR, semble expliqué la différence entre les établissements gagnants et perdants.

Par la suite, j'ai représenté l'analyse en PANEL, avec l'intervalle de confiance, de la complexité des patients, puis du taux d'adressage, afin de constater l'impact temporel, c'est-à-dire s'il y a eu des changements dans le temps. La valeur de référence est 0. Si le coefficient est positif, cela représente un impact positif dans le temps, sinon l'impact est négatif.

Analyse en PANEL sur la complexité des patients :

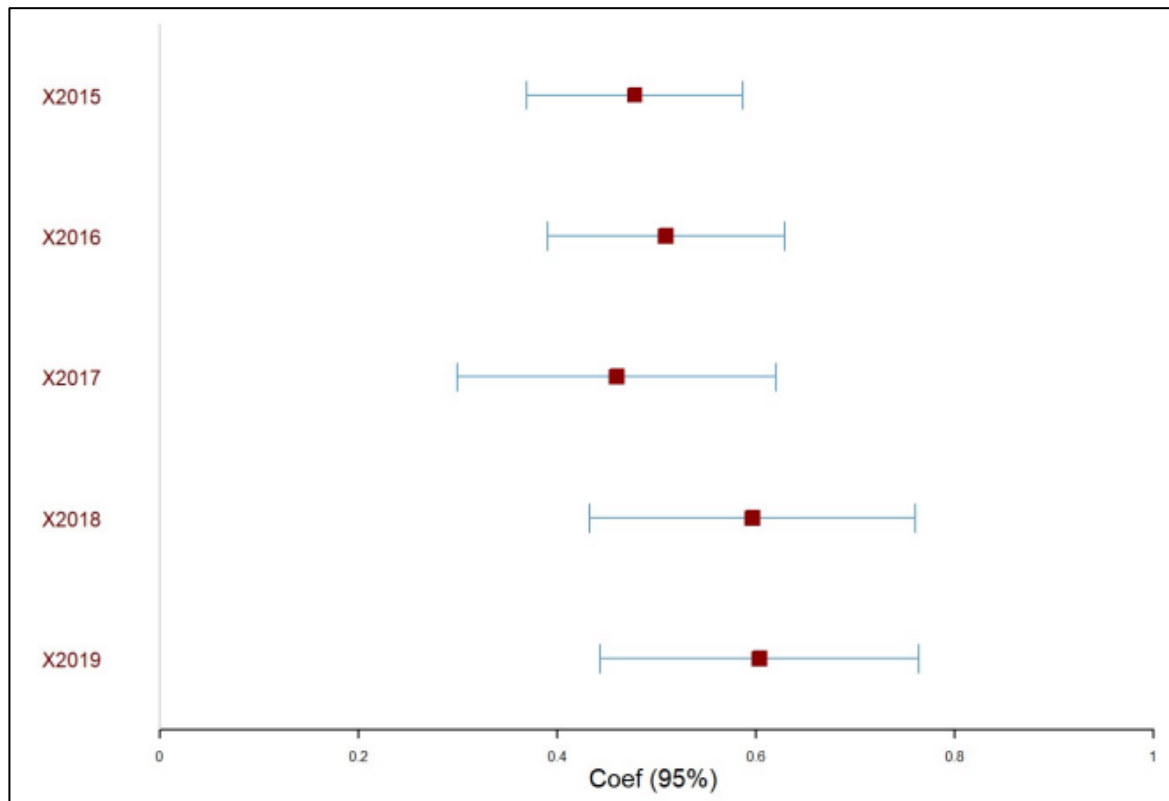


Tableau 11 : Coefficient de l'analyse en panel de la complexité des patients venant d'établissements MCO, vers des établissements SSR, pour les établissements de soins estimés gagnants

Interprétation : Depuis 2018, pour les établissements estimés gagnant, un patient ayant une lourde comorbidité à 60% (entre 43% et 77% selon l'intervalle de confiance) de chance de plus d'être transféré dans un établissement SSR, à la suite d'une opération orthopédique dans un établissement MCO, qu'un patient n'en possédant pas.

Analyse en PANEL sur le taux d'adressage :

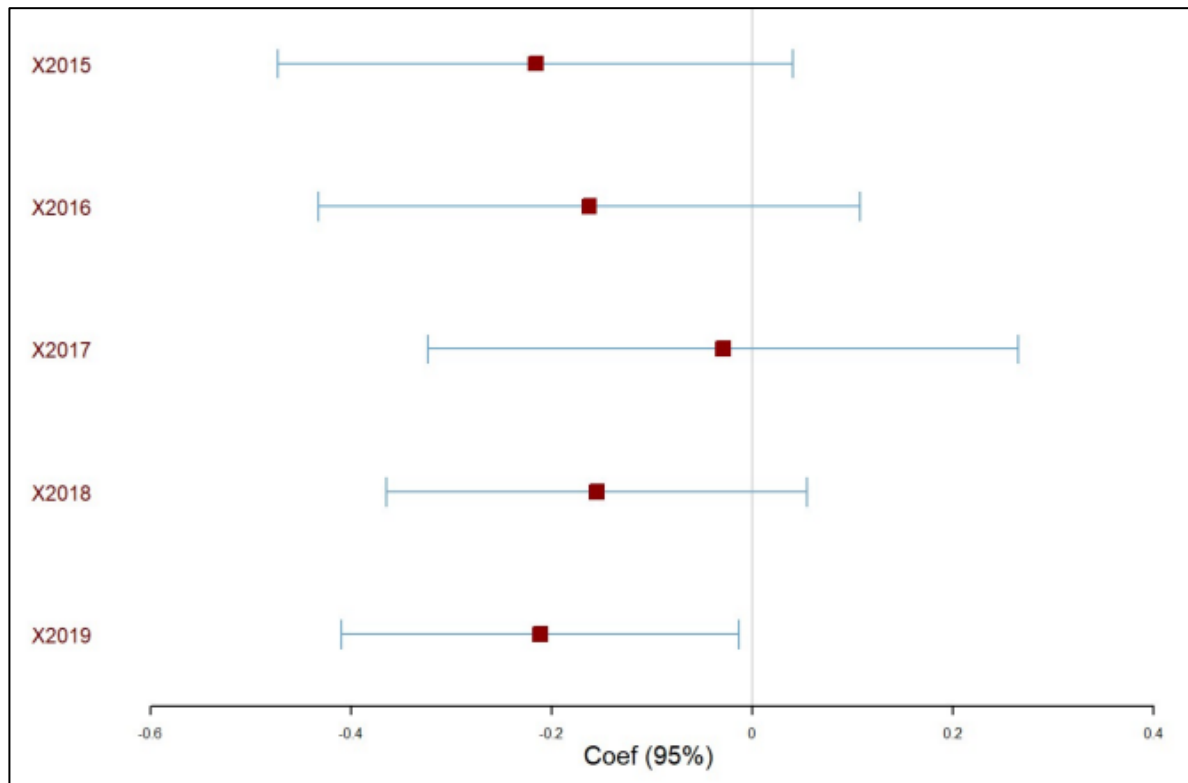


Tableau 12 : Coefficient de l'analyse en panel du taux d'adressage (%) de patient venant d'établissements MCO, vers des établissements SSR, pour les établissements de soins estimés gagnants

Interprétation : Depuis 2017, pour les établissements estimés gagnant, ce graphique semble indiquer que les taux d'adressage en SSR tendent à diminuer, et de façon significative en 2019 (l'intervalle de confiance étant inférieur à 0).

6.4 Conclusion

D'après les analyses univariées, déterminer que le forfait serait avantageux pour un hôpital, semble être lié au type de patient pris en charge, ainsi que son au taux d'adressage. De plus, la faible offre de soin par territoire, semble également être un élément déterminant dans ce contexte.

7 Conclusion

D'après nos analyses, il semblerait que les hôpitaux volontaires à l'expérimentation, semblent mieux ancrés et avoir plus d'activités dans leurs territoires, et donc mieux préparé à ce projet. De plus, il semblerait que la complexité des patients, ainsi que le taux d'adressage, définiraient si le forfait serait avantageux, ou non, pour les établissements, mais également lorsqu'ils possèdent un bon réseau de soin, et une certaine autonomie.

Cependant, concernant la détermination des établissements, dont le forfait semblerait avantageux ou non, notre modèle c'est concentré seulement sur les 30 établissements volontaires à l'expérimentation. La taille de cet échantillon peut-être un élément à prendre en comptes dans nos résultats obtenus.

8 Bilan personnel

Ce stage à l'Ecole des Hautes Etudes en Santé Publique fut très enrichissant pour moi. J'ai pu découvrir le domaine des statistiques dans le milieu professionnel, et y appliquer mes compétences apprises pendant ma formation.

J'y ai découvert quelques méthodes économétriques, et le secteur de la santé publique. Durant ce stage, j'ai également pu développer mon autonomie et mon organisation, et aussi mes connaissances statistiques.

De plus, j'ai fait de grand progrès sur le logiciel R, ce qui m'a permis d'avoir plus confiance en mes capacités sur ce logiciel.

9 Référence

1) RSMS :

<https://arenas.eu/recherche/axes-recherche/rsms/>

2) Présentation historique de l'EHESP

https://fr.wikipedia.org/wiki/%C3%89cole_des_hautes_%C3%A9tudes_en_sant%C3%A9_publique

3) Présentation de l'EHESP :

<https://www.ehesp.fr/ecole/missions-valeurs/>

4) Présentation du projet Episode De Soins :

(Voir dossier : Présentation_EDS)

5) Analyses bibliographiques :

(voir dossier : Analyse_Bibliographie)

10 Annexe

10.1 Programmation

Voici 3 codes, qui m'ont permis de réaliser mes analyses.

Comparaison des indicateurs réseaux, des établissements expérimentateurs et non-expérimentateurs (tableau 4) :

```
# Construction du tableau concernant l'analyse univarié
coef <- dataset3 %>%tbl_uvregression(method = glm,
                                   y = experimentateur,
                                   method.args = list(family = binomial(link="logit")),
                                   exponentiate = TRUE,
                                   hide_n = TRUE,
                                   conf.int = FALSE) %>% bold_p()
# On renomme les modalités de la variables experimentateur
dataset3 <- dataset3 %>% mutate(experimentateur = ifelse(experimentateur == "yes", "expe", "n-expe"))
# Construction du tableau concernant la comparaison des expe et n-expe
comparaison <- dataset3 %>% tbl_summary(by=experimentateur, digits = all_continuous() ~ 2) %>%
modify_header(label ~ "**Variable**") %>% bold_labels()
# Fusion des deux tableaux
tbl_comp <-tbl_merge(tbls = list(comparaison, coef),
                    tab_spanner = c("**Comparaison**", "**Modele univarie**"))

tbl_comp
```

Dans un premier temps, j'ai créé un premier tableau, à l'aide du package *gtsummary* et de la fonction *tbl_uvregression()*, qui permet de faire une analyse univarié. Ce modèle est une régression logistique, avec la variable *experimentateur* en Y. Puis, j'ai créé un deuxième tableaux, qui affiche la médiane, et le 1^{er} et 3^{ème} quartiles, des indicateurs réseaux des établissements expérimentateurs et non-expérimentateurs, grâce à la fonction *tbl_summary()*, également issu du package *gtsummary*. Enfin, j'ai fusionné ces deux tableaux, avec la fonction *tbl_merge()*.

La fonction *bold_p()*, permet de mettre en gras les pvalues significatifs, au seuil de risque 5%. La fonction *tab_spanner()*, permet de mettre un titre pour chacun des deux tableaux que j'ai fusionné. Enfin, la fonction *bold_labels()*, permet de mettre en gras le nom des variables.

Comparaison des indicateurs territoriaux, à partir de la simulation des établissements de soins estimés gagnants et perdants (tableau 10) :


```

{r, echo=FALSE, fig.align="center", warning=FALSE}
# Création du premier tableau (modèle multivarié)
data_mode <- data_univ[,c(-2:-6,-12)]
modell1 = glm(gagnant~., data=data_mode, family=binomial(link="logit"))
Anova(modell1, type="II", test="LR")
modell2 = stepAIC(modell1, direction = "both", k = 2, trace=FALSE)
Anova(modell2, type="II", test="LR")
modell2 %>% tbl_regression(exponentiate = TRUE, pvalue_fun = function(x) style_pvalue(x, digits = 3),
conf.int = FALSE) -> multi

# Création du deuxième tableau (univarié)
coef <- data_mode %>%tbl_uvregression(method = glm,
                                     y = gagnant,
                                     method.args = list(family = binomial(link="logit")),
                                     exponentiate = TRUE,
                                     hide_n = TRUE,
                                     conf.int = FALSE,
                                     pvalue_fun = function(x) style_pvalue(x, digits = 3)) %>% bold_p()
data_mode %>% mutate(gagnant = ifelse(gagnant==1, "winning", "losing")) -> data_mode

# Création du troisième tableau (Comparaison des indicateurs)
comparaison <- data_mode %>% tbl_summary(by=gagnant, digits = all_continuous() ~ 2) %>%
modify_header(label ~ "**Variable**") %>% bold_labels()

{r, echo=FALSE, warning=FALSE}
# Fusion des trois tableaux
tbl_comp <-tbl_merge(tbls = list(comparaison, coef, multi),
                    tab_spanner = c("**Comparison**", "**Univariate model**", "**Multivariate model**"))

tbl_comp

```

Dans un premier temps, j'ai effectué un modèle logistique complet, avec la variable *gagnant* en Y, puis j'ai sélectionné les variables, dont le modèle serait le plus performant selon le critère AIC. J'ai utilisé la fonction *stepAIC()*, puis j'ai créé un tableau, affichant l'Odds Ratio des variables, avec leurs pvalues, grâce à la fonction *tbl_regression()*. Ensuite, j'ai créé un deuxième tableau, affichant les résultats de l'analyse univarié. Puis j'ai créé un dernier tableau, qui affiche la médiane, et le 1^{er} et 3^{ème} quartile, des indicateurs territoriaux des établissements estimés gagnants et perdants. Enfin, j'ai fusionné ces trois tableaux.

Coefficient de l'analyse en panel de la complexité des patients venant d'établissements MCO, vers des établissements SSR, pour les établissements de soins estimés gagnants (tableau 11), coefficient de l'analyse en panel du taux d'adressage (%) de patient venant d'établissements MCO, vers des établissements SSR, pour les établissements de soins estimés gagnants (tableau 12) :

```

```{r, echo=FALSE, warning=FALSE}
Boucle permettant de faire une analyse en panel, année par année
coefplm_cpx <- c()
coefplm_gagnant <- c()
adress2 <- adress2 %>% drop_na()
for (i in 35:39){

 p <- plm(log(adress2[,i])~cpx+gagnant, data=adress2, index=c("finess"), model = "random")
 s <- summary(p)

 temp_cpx <- data.frame(p$coefficients[2], confint(p)[2,1], confint(p)[2,2], s$coefficients[2,4])
 temp_gagnant <- data.frame(p$coefficients[3], confint(p)[3,1], confint(p)[3,2], s$coefficients[3,4])
 coefplm_cpx <- rbind(coefplm_cpx, temp_cpx)
 coefplm_gagnant <- rbind(coefplm_gagnant, temp_gagnant)
 temp_cpx <- c()
 temp_gagnant <- c()
}

Les lignes prennent le nom des variables
rownames(coefplm_cpx) <- c("X2015", "X2016", "X2017", "X2018", "X2019")
rownames(coefplm_gagnant) <- c("X2015", "X2016", "X2017", "X2018", "X2019")

On nomme les 3 colonnes
colnames(coefplm_cpx) <- c("Coef", "Lower", "Upper", "p_value")
colnames(coefplm_gagnant) <- c("Coef", "Lower", "Upper", "p_value")

On transforme coefplm en data.frame
coefplm_cpx <- data.frame(coefplm_cpx)
coefplm_gagnant <- data.frame(coefplm_gagnant)
```

```{r, echo=FALSE, warning=FALSE}
Visualisation des coef avec leurs CI95% pour chaque variable
forestplot(rownames(coefplm_gagnant), coefplm_gagnant$Coef,
 coefplm_gagnant$Lower,
 coefplm_gagnant$Upper,
 zero = 0,
 xlab = "Coef (95%) ",
 col = fpColors(box = c("red4"),
 lines = c("skyblue3")),
 xticks = seq(-0.6, 0.4, 0.2),
 boxsize = 0.1,
 vertices = TRUE,
 txt_gp = fpTxtGp(label = list(gpar(fontfamily = "",
 col = "#660001", cex = .75)),
 ticks = gpar(fontfamily = "", cex = .5),
 xlab = gpar(fontfamily = "font", cex = .9)))
```

```

Dans un premier temps, j'ai créé 2 tableaux, pour les coefficients concernant la complexité des patients et du taux d'adressage des établissements estimés gagnants. Ensuite, j'ai effectué une boucle, allant de 35 à 39 (ce sont les positions des années 2015 à 2019 dans le data.frame), faisant une analyse en panel, de chacune des années, que j'ai enregistré dans la variable p, grâce à la fonction *plm()*. Ensuite, *temp_cpx* et *temp_gagnant* vont contenir les coefficients (complexité du patient pour *temp_cpx*, taux d'adressage pour *temp_gagnant*), avec leurs intervalles de confiance et leurs pvalue. Puis je les ajoute à *coefplm_cpx* et *coefplm_gagnant*. *rownames()* permet de nommer chaque ligne en fonction de l'année, et *colnames()* permet de nommer les variables.

J'ai utilisé la fonction *forestplot()*, afin de représenté graphiquement les coefficients, avec l'intervalles de confiances. Le code présenté, ci-dessus, concerne le taux d'adressage des établissements estimés gagnants.