



HAL
open science

Recours aux traitements de l'infertilité en France

Antoine Pointeaux

► **To cite this version:**

Antoine Pointeaux. Recours aux traitements de l'infertilité en France: Une étude géographique des déterminants contextuels. 2020. hal-03044372

HAL Id: hal-03044372

<https://hal.science/hal-03044372>

Preprint submitted on 7 Dec 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

DOCUMENTS DE TRAVAIL **260**

Recours aux traitements de l'infertilité en France :

Une étude géographique
des déterminants contextuels

Antoine Pointeaux

Antoine Pointeaux, *Recours aux traitements de l'infertilité en France : Une étude géographique des déterminants contextuels*, Paris, Ined, Document de travail, 260



Recours aux traitements de l'infertilité en France :

Une étude géographique des déterminants contextuels.

Auteur : *Antoine Pointeaux*

Résumé

Près de 10 % des femmes en âge de procréer dans le monde rencontrent un problème d'infertilité. Après 5 ans de rapports sexuels non protégés sans grossesse, seul un couple sur deux consulte un médecin en Europe. Cette étude s'intéresse aux déterminants contextuels du recours aux traitements de l'infertilité pour l'année 2016 en France métropolitaine. En s'appuyant sur le système national des données de santé (SNDS), cette étude écologique investigate le phénomène spatial et les disparités territoriales du taux de recours aux traitements de l'infertilité. L'échelle d'agrégation des données choisie pour mener l'étude est celle des territoires de vie (TV). Les facteurs étudiés sont la défavorisation sociale du TV (FDep13), l'accessibilité aux médecins généralistes et gynécologues (APL), le taux d'urbanisation, la distance au centre d'assistance médicale à la procréation (AMP) le plus proche et l'accès aux soins de santé de base. Pour prendre en compte l'hétérogénéité spatiale, une régression géographiquement pondérée (RGP) est utilisée. L'étude montre la présence d'autocorrélation spatiale et la nécessité de mettre en place des modèles géographiques. Elle met en évidence l'importance de l'accès aux médecins gynécologues dans le recours aux traitements de l'infertilité ainsi que l'effet équivoque de la favorisation sociale.

Mots-clés : infertilité, recours aux soins, autocorrélation spatiale, régression géographiquement pondérée, hétérogénéité spatiale

Abstract

About 10 % of women in reproductive age is infertile worldwide. In Europe, only 50 % of these couple seek for infertility service in a time frame of five years. This study focuses on contextual causes of infertility treatments use in 2016 in metropolitan France. Based on data from the national health data system this ecologic study investigates spatial phenomenon and local disparities in infertility treatment use. Analyzed factors are socioeconomic advantage index (FDep13), access in general practitioner and gynecologist (APL), urbanization rate, distance to the nearest assisted reproductive technology (ART) center and access to basic health care. In this study, the scale of the aggregation unit is the "territoire de vie" (TV). To explore spatial heterogeneity, we built a geographically weighted regression (GWR). This study shows the existence of spatial autocorrelation and the need of spatial models. It underlines the impact of gynecologist access in use of infertility treatments and the ambiguous impact of socioeconomic advantage.

Keywords: infertility health care use, spatial autocorrelation, geographically weighted regression, health care use, spatial heterogeneity

Liste des sigles et acronymes

Sigle ou acronyme	Signification
ACP	Analyse en composante principales
AIC	Critère d'information d'Akaike
AMP	Assistance médicale à la procréation
APL	Accessibilité potentielle localisée
AU	Aire urbaine
BV	Bassin de vie
CMDC	Cause médicale de décès
CMU-C	Couverture médicale universelle complémentaire
CV	Canton-ville
CV	<i>Cross validation</i>
DHHS	Department of Health and Human Services
EPCI	Etablissement public de coopération intercommunale
ESIS	European Study of Infertility and Subfertility
FIV (FIV-ICSI)	Fécondation <i>in vitro</i> (intra cytoplasmic sperm injection)
GLM	General linear model
GLMM	General linear mixed model
IA	Insémination artificielle
MAUP	Modifiable area unit problem
MCO	Moindre carré ordinaire
NSFG	National Survey for Family Growth
OMS	Organisation mondiale de la santé
PMSI	Programme de médicalisation des systèmes d'information
RGP	Régression géographiquement pondérée
SDM	Spatial Durbin Model
SEM	Spatial error model
SEIFA	Socio economic index for areas
SNDS	Système national des données de santé
SNIIRAM	Système national d'information inter-régimes de l'assurance maladie
SGLMM	Spatial General Linear Mixed Model
TV	Territoire de vie
UU	Unité urbaine
VIF	Variance inflation factor

Introduction

L'infertilité, un concept complexe

Le concept d'infertilité est complexe et non univoque. Sur le site de l'Organisation Mondiale de la Santé (OMS) (World Health Organization 2020) il est possible de trouver trois définitions différentes. Une première est la définition clinique de l'infertilité comme étant « l'incapacité de commencer une grossesse après 12 mois ou plus de rapports sexuels non protégés »¹. Une deuxième est la définition démographique comme étant « l'incapacité de celles en âge de se reproduire (15-49 ans) de commencer ou rester enceinte après 5 ans d'exposition »². Enfin une troisième définition épidémiologique pour des « femmes en âge de se reproduire (15-49 ans), à risque (ni enceintes ni allaitantes, sexuellement actives, n'utilisant pas de contraceptifs) qui reportent essayer sans succès de tomber enceinte depuis deux ans ou plus »³. Il existe donc différentes définitions qui peuvent prendre en compte l'âge de la femme, des critères de succès différents (être enceinte contre accoucher) ou des durées de la période de relations sexuelles non-protégées différentes (1 an, 2 ans, 5 ans). Il est également possible de différencier l'infertilité primaire pour les femmes nullipares de l'infertilité secondaire pour les femmes ayant déjà eu un enfant.

Un enjeu de santé publique

La multiplicité des définitions de l'infertilité rend difficile la comparaison des études estimant sa prévalence puisque les différents chiffres avancés par les équipes de recherche ne sont pas toujours comparables (et parfois les définitions retenues non-précisées). Selon l'OMS, l'infertilité touche plus de 10 % des couples dans le monde (World Health Organization 2020) pour des femmes « essayant d'avoir un enfant » et étant « dans une relation stable depuis cinq ans ». Dans une revue systématique regroupant au total 172 413 femmes à travers 25 pays, le chiffre de 9 % est avancé par Boivin et ses collègues en 2007 (Boivin, et al. 2007) pour estimer la prévalence de l'infertilité à 12 mois chez les femmes de 20 à 45 ans mariées ou dans une relation stable. Cela représente 72 millions de femmes à travers le monde en 2007 (entre 40 et 120 millions suivant les estimations « hautes » ou « basses »). D'après les statistiques de la *National Survey of Family Growth* (NSFG⁴), 8,8 % des femmes mariées âgées de 15 à 49 ans sont actuellement infertiles aux Etats-Unis. Ce chiffre couvre une large disparité selon la parité : 19 % pour les femmes n'ayant jamais eu d'enfant et 6,5 % pour celles ayant eu un enfant ou plus, toujours aux Etats-Unis. (DHHS 2019)

Des différences du Nord au Sud importantes

Au niveau mondial, l'impact de l'infertilité peut être très différent tant en termes de prévalence que de conséquences pour les femmes. En Afrique sub-saharienne, jusqu'à un tiers des femmes seraient touchées par l'infertilité au cours de leur vie, principalement en raison d'une infertilité secondaire très élevée (Daar et Merali 2001). Pour les femmes, « la stigmatisation peut être extrême dans les pays où les femmes infertiles sont vues comme des fardeaux pour le bien-être socioéconomique de la

¹ Traduit par l'auteur, texte original: "Defined by the failure to achieve a clinical pregnancy after 12 months or more of regular unprotected sexual intercourse".

² Traduit par l'auteur, texte original: "An inability of those of reproductive age (15-49 years) to become or remain pregnant within five years of exposure to pregnancy".

³ Traduit par l'auteur, texte original: "Women of reproductive age (15-49 years) at risk of becoming pregnant (not pregnant, sexually active, not using contraception and not lactating) who report trying unsuccessfully for a pregnancy for two years or more."

⁴ Encadré explicatif en page 4

communauté »⁵ (WHO 2010). En France, pays du Nord, les problèmes d'infertilité peuvent amener une culpabilité, une peur, un sentiment d'impuissance ou des difficultés maritales et économiques. (de La Rochebrochard, 2004; Daar et Mereli, 2001).

Taux de recours aux traitements de l'infertilité

Parmi les femmes infertiles, seulement 56 % ont recours à un traitement de l'infertilité dans les pays les plus développés, et 51 % dans les pays en développement (Boivin, et al. 2007). L'enquête *European Study of Infertility and Subfertility* (ESIS) de 1992 est cohérente avec cette estimation pour les pays européens avec 50 % des couples ayant consulté après quatre années d'infertilité (de La Rochebrochard 2004).

Aux Etats-Unis, selon la NSFG la proportion de femmes mariées âgées de 15 à 49 ans ayant eu recours au moins une fois à une assistance contre l'infertilité est de 12,7 %. Il est question ici d'assistance et non pas de traitements car ce chiffre prend en compte la demande de conseil (DHHS 2019). De plus, ce chiffre porte sur une population différente (femmes mariées uniquement) et une zone géographique différente (Etats-Unis contre Europe). Il n'est pas question ici de comparer les deux chiffres et les deux continents mais simplement de donner un ordre de grandeur au lecteur des différents taux de recours.

Le parcours de soin d'une femme

Lorsqu'un couple fait face à un problème d'infertilité, le diagnostic peut amener à soulever un problème d'origine masculine, féminine ou indéterminée (Ben Messaoud 2018). Quelle que soit l'origine (masculine, féminine ou inconnue), ce sont toujours les femmes qui sont traitées et cela explique pourquoi il n'est question que des femmes et des *acheteuses* de traitements dans ce document de travail. Les quelques lignes suivantes tentent de décrire le parcours de soin « standard » d'une femme. Il faut néanmoins être prudent, les parcours de soins dépendent de bilans d'infécondité et les choix de traitements peuvent varier en fonction de ces bilans, du praticien et des souhaits et caractéristiques du couple.

Dans les faits, le premier traitement proposé semble souvent être de la famille des anti-œstrogènes (citrate de clomifène). Il peut être prescrit par un médecin généraliste ou un médecin gynécologue. En théorie, il ne devrait être prescrit que dans certains cas (en fonction du bilan d'infécondité) mais en pratique il est souvent utilisé comme traitement de première intention. En cas d'échec, les traitements de la famille des gonadotrophines sont utilisés. Les gonadotrophines ne peuvent pas être prescrits par un médecin généraliste. Ils sont utilisés à plusieurs fins : soit pour une stimulation hormonale simple, soit dans le but de préparer une insémination artificielle (IA) ou une fécondation *in vitro* (FIV, FIV-ICSI) (Ben Messaoud 2018). Toutes les femmes qui ont recours à une assistance médicale à la procréation (AMP) achètent des gonadotrophines.

Cette étude s'intéresse aux acheteuses de traitements d'infertilité (anti-œstrogènes et gonadotrophines), ce qui englobe également les femmes traitées par AMP. Elle concerne donc toutes les femmes traitées pour l'infertilité, quel que soit le traitement ou la technique. La littérature existante se concentre très majoritairement sur l'AMP en oubliant parfois que cela reste « le traitement de dernier recours » (Ben Messaoud 2018). L'une des originalités de cette étude est de prendre en compte également les traitements de premières intentions.

⁵ Traduit par l'auteur, texte original: "The stigmatization can be extreme in some countries where infertile people are viewed as a burden on the socioeconomic well-being of a community". Bulletin WHO, 2010; 88:881-882.

Bien que l'étude du parcours médical des femmes face à l'infertilité représente une part importante de la compréhension du phénomène de recours aux traitements, il est important de préciser ici que cette étude ne porte pas sur les femmes mais sur leurs environnements. C'est une étude écologique où les individus statistiques sont les territoires de vie des femmes.

Problématique de l'étude

L'ambition de cette étude est donc d'explorer l'accès à l'ensemble des traitements de l'infertilité sur le territoire de France métropolitaine à partir de l'analyse des données du système national des données de santé (SNDS). Les données du SNDS sont des données de remboursements et elles sont mobilisables dans le cadre de cette étude car les traitements de l'infertilité sont remboursés à 100 % pour les stimulations hormonales et jusqu'à 4 FIV et 6 IA pour les actes d'AMP (Ameli 2010). À défaut de pouvoir mobiliser des données individuelles sur les femmes, il s'agit de **chercher les déterminants contextuels à l'accès aux traitements de l'infertilité pour les femmes en France métropolitaine**. Tout au long de l'étude, l'accent sera mis sur **l'analyse spatiale de ce phénomène**. Notamment, cette étude cherche à investiguer deux hypothèses : l'absence d'autocorrélation spatiale et la stationnarité du phénomène spatial (s'il existe).

Partie I : Revue de littérature

Des inégalités durant les soins dans les centres d'AMP

Une première série d'études réalisées aux Etats-Unis constatent des inégalités durant les soins dans les centres AMP. Les inégalités observées concernent les revenus, les niveaux de formation et l'ethnicité des femmes. Smith et al s'appuient sur une étude incluant 391 femmes américaines dans des centres de traitements de l'infertilité pour montrer que les couples avec un haut niveau de diplôme et de revenus dépensent plus d'argent dans les cycles de traitements et augmentent leurs chances de grossesse (Smith, et al. 2011). En effet, ces couples à hauts revenus ont des parcours plus longs (ils réalisent plus de cycles de traitements) et des traitements plus performants (insémination artificielle et fécondation in vitro).

De façon analogue, sur 1 500 femmes se présentant dans ces centres d'AMP, Jain et al observent que les femmes afro-américaines ont des durées d'infertilité plus longues avant de recourir à des traitements de l'infertilité que les femmes caucasiennes (Jain 2006). Les femmes afro-américaines ont des durées moyennes d'infertilité de 4,5 ans contre 3,5 ans pour les femmes blanches. Cette différence pourrait s'expliquer par le niveau de diplôme et de revenus des femmes. En effet, les femmes afro-américaines et hispaniques ont généralement des niveaux de diplômes moins élevés et des revenus plus faibles que les femmes caucasiennes dans ces centres AMP.

Ces études souffrent d'un biais de sélection car les femmes y sont incluses sont celles qui se rendent dans les centres AMP. Dès lors, seules les femmes ayant recours à ces traitements d'AMP sont observées. Elles n'incluent pas les femmes n'ayant pas recours à ces traitements, c'est-à-dire les femmes traitées médicalement pour l'infertilité sans AMP, et il n'y a pas d'informations sur l'ensemble des femmes concernées (femmes en âge reproductif). Cependant, ces études ont l'intérêt de constater les inégalités dans les centres d'AMP. Elles semblent montrer que de nombreuses barrières existent pour les femmes, des barrières d'éducation, de revenus ou de culture.

Des inégalités dans l'accès aux soins de l'infertilité : une première barrière culturelle

Après avoir constaté une série d'inégalités de revenus, de diplômes et d'ethnicité chez les femmes dans les centres AMP, une seconde série d'articles se penchent sur les barrières potentielles à l'accès aux soins des traitements de l'infertilité, principalement aux États-Unis et en Australie.

Le National Survey for Family Growth (NSFG)

Le NSFG est une enquête menée au Etats-Unis et conduite par le National Center for Health Statistics (NCHS) qui a pour objectif de rassembler des informations et de produire des statistiques sur la vie familiale dans le pays : mariages, divorces, natalité, infertilité, contraception. La première enquête fut menée en 1973 pour être représentative des femmes mariées âgées de 15 à 44 ans aux Etats-Unis. En 2002, les hommes furent rajoutés à l'enquête et en 2015 la tranche d'âge fut élargie de 15 à 49 ans. Au départ sous forme de cycles (premier cycle en 1973, second en 1976, troisième en 1982 etc), le NSFG est produit sous forme continue depuis 2006. Les entretiens sont menés tout au long de l'année (quatre période de 12 semaines chaque année). Les entretiens sont menés pas des femmes, la partie la plus sensible du questionnaire étant auto-administrée. L'enquête est volontaire et confidentielle, le taux de réponse pour les données les plus récentes est d'environ 70 %. (DHHS 2019)

La première barrière est une barrière ethnique et culturelle. En s'appuyant sur les données du NSFG sur la période 2006-2010 aux Etats-Unis, Chandra et al soulignent qu'avoir eu accès aux traitements de l'infertilité est plus fréquent pour les femmes blanches (15 %) que pour les femmes hispaniques (7,6 %) ou afro-américaines (8 %) (Chandra, Copen et Stephen 2014). Selon Quinn et Fujimoto, ces différences ethniques sont à relativiser car les données portent sur les femmes mariées. Or, « étant donné les différences de prévalence des naissances hors mariage en fonction des ethnicités, les études sur les données du NSFG introduisent un biais de sélection et la validité des résultats doit être remise en cause »⁶ (Quinn et Fujimoto 2016). Néanmoins, pour expliquer ces différences de recours aux traitements, Adashi et al soulignent que « les barrières socioculturelles semblent jouer un rôle majeur dans le manque d'accès aux soins des traitements de l'infertilité »⁷ (Adashi et Dean 2016). White et al soulignent que ces différences peuvent être dû à « une aversion culturelle pour les solutions technologiques aux problèmes de santé, [...] une plus grande méfiance vis à vis d'un système dont ils étaient victimes dans le passé »⁸ (White, McQuillan et Greil 2006). Quinn et al restent plus prudent et déclarent « qu'un des facteurs pourrait être une stigmatisation plus forte des femmes ayant besoin de ces soins parmi certaines communautés »⁹ (Quinn et Fujimoto 2016).

La barrière économique : la plus discriminante ?

La seconde barrière est la barrière économique et financière. Selon Quinn, « le prix des traitements est la plus grande barrière à l'accès aux traitements de l'infertilité aux Etats-Unis »¹⁰ (Quinn et Fujimoto 2016). Le faible taux de remboursement de ces traitements aux Etats-Unis provoque des coûts très élevés qui rendent la situation « socialement injuste : le droit de fonder une famille face aux problèmes d'infertilité est désormais lié au niveau de revenus »¹¹ (Adashi et Dean 2016).

Si la situation aux Etats-Unis est inégalitaire au regard des coûts supportés par les couples en raison de remboursements faibles, cette situation pourrait être différente dans les autres pays ayant des politiques de santé différentes. L'article de Harris et al étudie cette question en Australie où « les traitements de l'infertilité ont historiquement été subventionnés par le système d'assurance maladie national, Medicare »¹² (Harris, et al. 2016). Malgré cela, ils montrent que les femmes habitant dans des territoires socioéconomiquement désavantagés ont moins accès aux traitements que les femmes habitant dans des zones plus favorisées. L'intuition serait de conclure que les couples vivant dans ces zones avantagées ont plus de revenus pour financer les cycles de traitements, car bien que subventionnés, le coût d'un cycle de transfert d'embryon supporté par un couple australien varie de

⁶ Traduit par l'auteur, texte original: "Given ethnic variances in prevalence of non-marital birth, the NSFG study introduces ethnic bias, and the validity of its conclusions is called into question » (p.1120)

⁷ Traduit par l'auteur, texte original: "sociocultural barriers to access to care appear to play a major role in compromising the availability of general infertility and ART services » (p.1114-1115)

⁸ Traduit par l'auteur. Texte original: "cultural aversion to technological solutions to a holistic health problem, greater aversion to infertility treatments and labels, greater distrust of a medical establishment that has victimized minorities in the past » (p.856)

⁹ Traduit par l'auteur, texte original: "Some of these factors may include concern regarding the stigma of infertility delaying initiation of treatment in some ethnic and racial groups » (p.1121)

¹⁰ Traduit par l'auteur, texte original: "The cost of care has been identified as the greatest barrier to access to infertility care in the US » (p.1120)

¹¹ Traduit par l'auteur, texte original: "This state of affairs is socially unjust: the right to build a family in the face of infertility is now tied to economic prowess » (p.1117)

¹² Traduit par l'auteur, texte original: "In Australia, assisted reproductive technology treatment has historically been subsidised through the public health insurance scheme, Medicare » (p.576)

1 000 à 2 700 euros selon le type de transfert¹³. Cette situation est difficilement comparable avec la France où les soins sont remboursés à 100 %. Cependant, une étude menée aux Etats-Unis, où le coût supporté par le couple est supérieur à la situation australienne, montre que les couples arrêtent le traitement principalement pour des raisons psychologiques (Domar, et al. 2010). Il existe au moins deux autres hypothèses pour expliquer que les différences entre groupes socioéconomiques se maintiennent malgré une prise en charge financière. La première hypothèse est une différence de comportement vis-à-vis des recours aux soins entre les différents groupes socioéconomiques indépendamment de la politique de remboursement. La seconde hypothèse est une identification différente des problèmes d'infertilité entre ces groupes socioéconomiques. En effet, une étude de Morris et al en 2010 portant sur une population anglaise (n=1 486) montre que les femmes appartenant à un groupe socioéconomique plus désavantagé déclarent moins de problème d'infertilité que celles des groupes socioéconomiques plus avantagés. Cette étude contrôle l'âge de la femme lorsqu'elle essaie pour la première fois de tomber enceinte. L'interprétation de ce résultat n'est pas de conclure à une différence de fertilité entre les groupes socioéconomiques mais plutôt à une différence d'identification de ces problèmes d'infertilité (Morris, et al. 2011).

La distance aux infrastructures

Une autre barrière est celle de l'éloignement aux infrastructures de santé. Les femmes australiennes vivant dans les grandes villes ont plus accès aux soins que celles vivant dans des zones moins urbanisées (Harris, et al. 2016). Aux Etats-Unis, « l'accès physique peut être un problème au vu de l'hétérogénéité de la distribution des gynécologues et des centres de reproduction sur le territoire » (Davis et Sokol 2016). Une étude environnementale sur les cinquante états des Etats-Unis montre que le taux de recours est corrélé avec le nombre de praticiens et de centres de l'état, ainsi qu'avec le revenu médian de l'état et son taux d'assurance maladie (Hammoud, et al. 2009). Malheureusement, ce niveau géographique de l'état ne permet pas une analyse plus fine sur les zones urbaines ou rurales.

Conclusion de l'état de la littérature et motivation de cette étude

L'infertilité est un problème de santé majeur qui touche au minimum 70 millions de couples à travers le monde. Parmi ces couples, seulement un sur deux ont recours aux traitements de l'infertilité. Plusieurs inégalités sont observées dans les parcours de soins et dans l'accès aux traitements de l'infertilité. Les barrières à cet accès peuvent être d'ordre psychologiques, culturelles, économiques, géographiques et d'accès aux infrastructures. La recherche, principalement concentrée aux Etats-Unis et en Australie, reste rare sur l'accès aux traitements de l'infertilité. De l'aveu de Harris et al « des études sont nécessaires pour comprendre les différences d'infertilité entre les groupes socioéconomiques et les moyens efficaces de lutter contre les inégalités d'accès à ces soins »¹⁴ (Harris, et al. 2016). La recherche est même quasiment inexistante en France pour le moment, et très rare au niveau européen. De plus, les études se concentrent sur les techniques d'AMP (insémination artificielle, fécondation in vitro). Ainsi, il semble que des études spécifiques sur tous les traitements de l'infertilité soient nécessaires.

¹³ L'article mentionne entre 3 000 à 4 000 dollars australiens pour un nouveau transfert d'embryon et entre 1 500 et 2 000 dollars australiens pour un transfert d'embryon congelé. Le taux de change utilisé est le taux de change au 31 décembre 2015 (1 AUD = 0,67 E)

¹⁴ Traduit par l'auteur, texte original: "studies are needed to fully understand differences in the burden of subfertility among different socioeconomic groups and effective ways to ameliorate inequalities in access to care » (p.583)

Partie II : Matériel

2.1 Les données du SNDS et la mesure du taux de recours

Cette étude s'appuie sur les données du système national de données de santé (SNDS). Créé en 2016, le SNDS rassemble plusieurs bases de données, dont celles du système national d'informations inter-régimes de l'assurance maladie (SNIIRAM), créé en 1998. Le Sniiram centralisait déjà depuis 1998 les données de l'Assurance maladie afin de contribuer à une meilleure gestion de l'assurance maladie et des politiques de santé. La loi de modernisation du système de santé français de janvier 2016 élargit le dispositif avec la création du SNDS. L'objectif est double : d'une part, augmenter la connaissance des dépenses de santé, l'information sur les professionnels de santé, les structures et leurs activités. D'autre part, ouvrir les données à la recherche, aux études et aux évaluations des politiques de santé. Le SNDS est aujourd'hui composé :

- Du Sniiram ;
- Des données des établissements de santé (Programme de Médicalisation des Systèmes d'Information - PMSI) ;
- Des données sur les causes médicales de décès (CMDC)

L'utilisation du SNDS est possible dans cette étude car les traitements de l'infertilité sont remboursés à 100 %. L'accès aux soins de l'infertilité est mesuré par le taux de recours aux médicaments anti-œstrogènes et gonadotrophines au niveau de l'unité géographique analysée¹⁵. Ces médicaments sont utilisés en première intention mais également pour réaliser une AMP. Ces données couvrent donc toutes les femmes infertiles ayant recours aux traitements de l'infertilités. Le taux de recours est le rapport entre le **nombre d'acheteuses de traitements âgées de 18 à 49 ans durant l'année 2016** au niveau de l'unité géographique analysée et le **nombre total de femmes âgées de 18 et 49 ans** au niveau de l'unité géographique analysée. Les données exhaustives n'étant pour le moment disponible que jusqu'en 2016, l'étude se concentre sur le taux de recours calculé en 2016.

Le nombre d'acheteuses de traitements dans le territoire de vie est obtenu à l'aide du système national de données de santé (SNDS) au niveau communal. Pour certaines communes, il n'a pas été possible de récupérer l'information. Cela fut le cas dans trois cas de figure :

- Le code commune est erroné. Cela représente 334 communes et 1 929 acheteuses de traitements.
- Le code commune est périmé. Cela représente 583 communes et 2 048 acheteuses de traitements.
- Pour des raisons de confidentialité, les codes commune pour lesquelles l'effectif de femmes âgées entre 18 et 49 ans est inférieur à 5 ne sont pas récupérés. Cela représente 13 communes et 14 acheteuses de traitements.

Le nombre de femmes âgées de 18 à 49 ans est une donnée disponible au niveau communal en libre accès sur le site [data.gouv.fr](https://www.data.gouv.fr/fr/datasets/population-municipale-t-popmun-com/#_) à l'adresse suivante : https://www.data.gouv.fr/fr/datasets/population-municipale-t-popmun-com/#_. Ces données sont celles du dernier recensement disponible de 2016.

Cet indicateur mesure le recours aux soins de l'infertilité durant l'année 2016 mais n'est pas une mesure de la prévalence de l'infertilité. En effet, il ne prend pas en compte les femmes infertiles n'ayant pas eu recours à ces traitements.

¹⁵ Le niveau de l'analyse géographique utilisée est explicité après dans la partie « Choix de l'échelle d'agrégation spatiale »

2.2 Les variables explicatives de l'étude

La revue de littérature a permis de faire émerger des hypothèses sur les différentes barrières à l'accès aux traitements de l'infertilité. Ces barrières peuvent être de différentes natures, mais il semble s'agir notamment de barrières culturelles, psychologiques, économiques, géographiques et d'accès aux infrastructures. Chaque variable présentée ci-dessous cherche à mesurer une de ces barrières potentielles. Les barrières culturelles et psychologiques n'ont pas pu être prises en compte dans cette étude écologique. En particulier, la barrière psychologique est nécessairement individuelle et non pas contextuelle.

Le Tableau 1 présente les variables explicatives de l'étude et la barrière à l'accès aux traitements de l'infertilité qu'elles ambitionnent d'approcher. Pour les besoins de l'étude, les variables quantitatives de défavorisation sociale (Rey, Jouglu, et al. 2009), d'accessibilité aux médecins gynécologues et généralistes (Vergier et Lefebvre-Hoang, Déserts médicaux : Comment les définir ? Comment les mesurer ? 2017) et d'urbanisation ont par la suite été transformées en variables catégorielles par quantiles. Dans un premier temps elles sont découpées par quintiles puis par terciles pour la régression géographiquement pondérée (RGP). De plus amples informations sur les variables explicatives sont disponibles en Annexe A : Glossaire des variables explicatives.

Tableau 1 : Présentation synthétique des variables explicatives

Barrière mesurée	Nom de la variable	Description	Unité
Economique	FDep13	Indice de défavorisation sociale	
Offre de soin	APL_GEN	Accessibilité potentielle localisée pour les médecins généralistes	Nombre de consultations par habitant et par an (C/hab/an)
	APL_GYN	Accessibilité potentielle localisée pour les médecins gynécologues	C/hab/an
	Acces_sante	Accès aux services de santé de base en moins de dix minutes	Catégorielle
Ruralité/Urbanisme	Urban_tv	Taux d'urbanisation	Proportion
Infrastructure	DIST_AMP	Distance au centre AMP le plus proche	Catégorielle
	ACT_AMP	Activités des centres AMP	Nombre de ponctions d'ovocytes réalisées dans le centre

Source: Dares, Insee, Inserm, Ined

2.3 Choix de l'échelle d'agrégation spatiale

Notre objectif est de mener une étude écologique si bien que nous devons construire une base de données dont le niveau d'observation est une unité géographique. Plusieurs niveaux géographiques étaient envisageables *a priori* : communes, départements, niveaux intermédiaires, etc. La seule information à disposition pour les acheteuses de traitements d'infertilité est la commune de résidence. Leur catégorie socioprofessionnelle, leur revenu ou le nombre de consultations chez un médecin gynécologue sont indisponibles. Avec un accès au SNDS, il aurait été possible de savoir si la femme était bénéficiaire ou non de la CMU-C ainsi que son âge. La crise pandémique n'a pas permis cet accès au SNDS. Cependant, il est possible de qualifier l'environnement dans lequel la femme réside. L'hypothèse émise est de dire que cet environnement va avoir un impact sur le comportement des femmes et, in fine, sur le taux de recours à ces traitements de l'infertilité dans cet environnement. Dans une étude écologique, il n'est pas possible de conclure sur les déterminants individuels, cela serait commettre l'**erreur écologique**. Dès lors, le choix de l'échelle spatiale utilisée a une importance particulière. D'une part, ce niveau de découpage du territoire forme les individus statistiques de l'étude. D'autre part, ce découpage impacte directement l'analyse et l'interprétation des résultats. Il doit correspondre au mieux à une réalité « de vie » pour les populations le composant vis-à-vis de la variable d'intérêt.

La nécessité d'une échelle intermédiaire

Une première échelle accessible (dans le sens où les données sont accessibles à cette échelle) est le département. Cependant, choisir pour l'étude une échelle aussi vaste risque de masquer des inégalités d'accès au sein du département. Notamment, l'échelle départementale ne permet pas de faire la distinction entre une zone plus rurale et une zone plus urbaine. Pour que le travail soit pertinent, il semble nécessaire de travailler à une échelle plus fine.

Une seconde échelle, plus fine, serait l'échelle communale¹⁶. Les différentes variables utilisées dans l'étude sont disponibles à ce niveau, mais cette échelle s'est révélée trop fine. En effet, plus d'un tiers des communes de France ont moins de 300 habitants, deux tiers moins de 800 habitants. Dès lors, des indicateurs comme le nombre de femmes ayant eu accès aux traitements de l'infertilité au cours d'une année ou encore le profil des gynécologues de la commune n'ont pas beaucoup de sens car les effectifs sont trop faibles.

Le choix de l'échelle intermédiaire

Il faut donc trouver un équilibre entre une échelle trop fine qui risque de rendre certaines variables inutilisables et une échelle trop grande qui risque de masquer les effets que l'étude souhaite mettre en avant. L'Insee propose et utilise différentes échelles intermédiaires. Parmi elles, les EPCI (établissements publics de coopérations intercommunales), les bassins de vie, les territoires de vie, les territoires de vie-santé, les cantons villes par exemple. Le présente le nombre de zones correspondant à ces différentes échelles intermédiaires sur le territoire de France métropolitaine.

¹⁶ L'échelle la plus fine serait l'IRIS mais les données du SNDS ne sont pas disponibles à cette échelle.

Tableau 2 : Nombre d'individus statistiques créés par chaque échelle intermédiaire

Niveau	EPCI	CV	UU	AU	BV2012	TV
Nbre de zones en France métropolitaine	1255	2066	2304	771	1654	2677

Note de lecture : La France comporte 771 aires urbaines.

- EPCI = établissements publics de coopérations intercommunales
- CV = canton ville
- UU = unité urbaine
- AU = aire urbaine
- BV2012 = bassin de vie 2012
- TV = territoire de vie

Source : Insee

Si les aires urbaines semblent trop peu nombreuses et risquent également de masquer des effets, les autres découpages peuvent être envisagés. Notamment les bassins de vie et territoires de vie, dont les définitions semblent le mieux correspondre au découpage voulu, c'est-à-dire des unités géographiques pouvant caractériser l'accès aux soins des populations y habitant. En effet, « ils sont définis comme les plus petits territoires, organisés autour d'un pôle de services, au sein desquels la population a accès aux équipements et services les plus courants. Dans les contours d'un bassin de vie s'organise une grande partie du quotidien des habitants » (Reynard et Vialette 2014). Le niveau du territoire de vie est préféré car il découpe les bassins de vie de plus de 50 000 habitants. Il a été décidé pour cette étude de découper les communes de Paris, Lyon et Marseille selon leurs arrondissements afin d'éviter des individus statistiques regroupant trop d'habitants.

2.4 Exploration des données

2.4.1 Méthodes utilisées pour l'exploration des données

La première étape dans l'analyse des données est de réaliser une série de statistiques descriptives afin d'explorer les données. Ces statistiques exploratoires prennent plusieurs formes : des statistiques univariées et des cartes pour analyser les variables de l'étude une à une. Les cartes permettent de visualiser les différents phénomènes (accessibilité des médecins, défavorisation sociale etc) et de commencer l'approche géographique de l'étude. Dans un second temps, des statistiques bivariées et le tableau des corrélations permettent de se rendre compte des liens qui peuvent exister entre les variables. Enfin, une analyse en composantes principales (ACP) est réalisée avec deux objectifs. Le premier objectif de l'ACP est de mesurer l'inertie des axes construits. En effet, l'ACP vise à réduire le nombre de dimensions d'une analyse en combinant linéairement les variables pour former des composantes principales, ou des axes. Les axes sont sélectionnés grâce à la quantité d'information (inertie) qu'ils restituent. Le second objectif de l'ACP est de représenter graphiquement les variables en les projetant sur le sous-espace construit par les composantes principales afin de mieux comprendre les liens qui peuvent exister. Ces deux objectifs serviront également à mesurer la complexité du phénomène étudié, c'est-à-dire le recours aux traitements de l'infertilité. En effet, la pertinence d'un modèle plus complexe comme la régression géographiquement pondérée (RGP) dépend de la complexité du phénomène et des données de l'étude.

2.4.2 Résultats de l'exploration des données

Une variabilité du taux de recours aux traitements de l'infertilité en France métropolitaine

Le Tableau 3 montre les premières statistiques descriptives de l'étude. Le taux de recours pour mille femmes varie de 0 à 60 selon le territoire de vie avec une moyenne de 9,94 pour mille femmes. Le premier quartile est à 7,72 et le troisième quartile à 11,87. Tableau 3 : Statistiques descriptives des variables de l'étude. La Figure 1 représente la carte des taux de recours aux traitements de l'infertilité dans les différents territoires de vie en France. Cette carte montre visuellement une variabilité du phénomène. Les taux de recours sont élevés le long de la côte atlantique, de la façade méditerranéenne, dans la région de Lyon ou encore en Ile-De-France. Une grande zone de taux de recours élevés se situe également à l'est de la région parisienne, dans le nord de la Champagne-Ardenne. Au contraire, les territoires de vie de l'Occitanie, de la Bourgogne-Franche-Comté, du sud du Centre Val-De-Loire ou de l'intérieur de la Nouvelle-Aquitaine ont des taux de recours plus faibles. L'intérêt de visualiser ces variables est de **constater une variabilité sur le territoire français**. Cela n'est pas suffisant pour assurer la pertinence des modèles spatiaux mais valide l'idée d'explorer cette piste pour comprendre le recours aux traitements de l'infertilité.

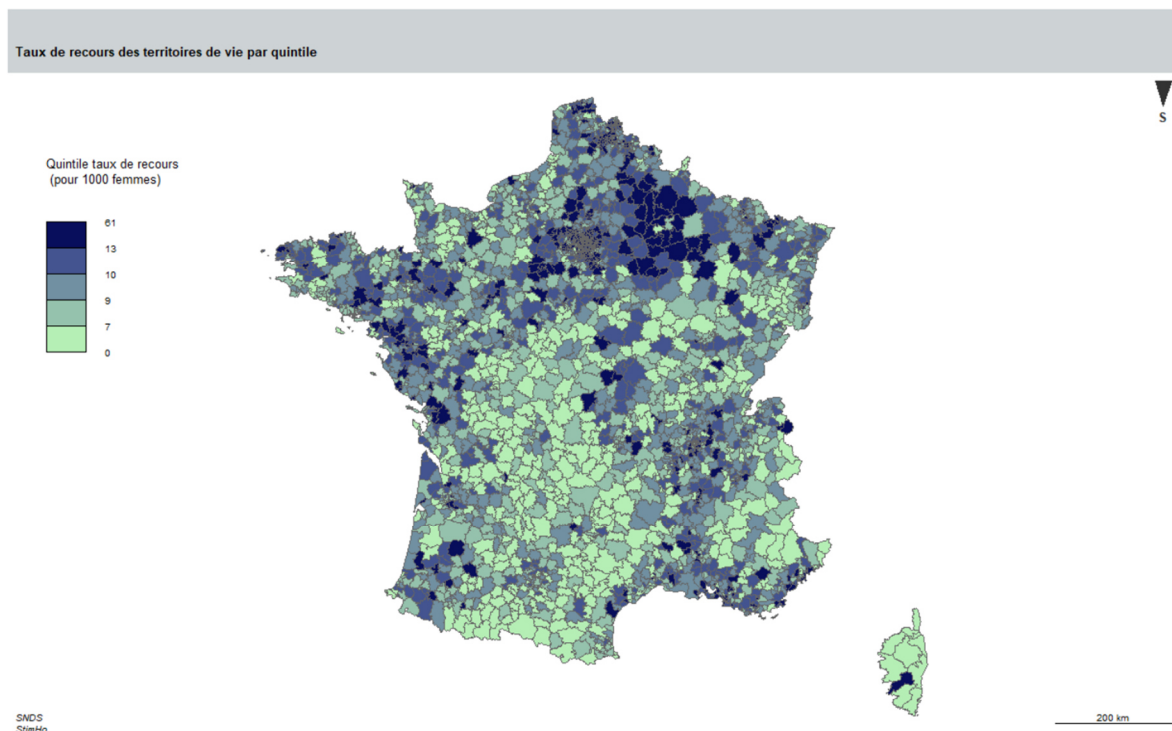


Figure 1 : Taux de recours pour mille femmes par territoire de vie en France métropolitaine

Une variabilité observée sur les autres variables de l'étude

Le Tableau 3 présente également les statistiques descriptives des autres variables de l'étude. Pour les médecins généralistes, le nombre moyen de consultations disponibles par an et par habitant est d'environ 4. Pour les médecins gynécologues, ce nombre de consultations disponibles est d'environ 5.

Cependant, la variabilité est plus grande pour les médecins gynécologues car certains territoires de vie n'ont aucune consultations disponibles (le minimum est 0) alors que le minimum pour les médecins généralistes est de 1,5 C/an/hab. L'écart entre le troisième et le premier quartile est plus élevé pour les médecins gynécologues que pour les médecins généralistes (4,57 contre 1,3). La variabilité du taux d'urbanisation est très élevée car il existe des territoires de vie dans lesquels la proportion de surface urbanisée est de 0 % et des territoires de vie où cette proportion est de 100 %. La différence entre la médiane (8,7 %) et la moyenne (22,2 %) indique que la distribution est asymétrique avec des valeurs extrêmes hautes. Autrement dit, il existe un nombre limité de territoire de vie très urbanisés. Cependant, le Tableau 8 (Annexe) montre que 40 % de la population vit dans le quintile des territoires de vie les plus urbanisés lorsque seulement 8 % vivent dans le quintile des territoires de vie les moins urbanisés. La variable *accès santé* présente beaucoup moins de variabilité. La médiane est à 100 % et le premier quartile à environ 97 %. Cependant le minimum est à 48 %, ce qui laisse penser que certains territoires de vie sont largement désavantagés du point de vue de cette variable.

Ces variables sont présentées visuellement en annexe (voir Annexe B : Les cartes des variables FDep, APL, Urbanisation, Activités des centres et accès santé). Pour l'indice de défavorisation sociale, l'influence des métropoles (Paris, Rennes, Nantes, Lille, Strasbourg, Dijon, Lyon, Marseille, Toulouse, Bordeaux, Reims, Orléans, Rouen, Poitiers) est très marquée. Les territoires de vie en Savoie et Haute-Savoie, proche de la ville de Genève en Suisse ont également des situations de favorisation sociale. Le corolaire de cette favorisation des grandes villes est la défavorisation sociale des zones rurales. Pour l'accessibilité aux médecins généralistes certaines régions semblent défavorisées, comme l'Ile-De-France, le Centre Val-de-Loire ou la Bourgogne Franche-Comté. Les Haut-De-France et le sud de la France semblent avoir un APL pour les médecins généralistes élevé, tout comme la côte atlantique. Les territoires de vie ayant un APL gynécologue élevé sont plus éparses sur les territoires mais certaines régions semblent se démarquer. Ces régions sont l'Ile-De-France, l'Alsace, et les alentours de Lyon, Marseille, Toulouse, Bordeaux ou encore Nantes et Lille. Ici de nouveau, l'influence des grandes villes semble jouer et cela alerte sur une possible colinéarité entre l'urbanisation, la défavorisation sociale et l'accessibilité aux médecins gynécologues.

Tableau 3 : Statistiques descriptives des variables de l'étude

	Min	1st quartile	Médiane	Moyenne	3e quartile	Max
tx_recours (pour mille femmes)	0	7.72	9.64	9.94	11.87	60.54
fdep13	-3.72	-0.40	0.22	0.08	0.66	2.33
apl_gyn (C/hab/an)	0	2.44	4.58	5.10	7.01	21.22
apl_gen (C/hab/an)	1.5	3.2	3.8	3.92	4.5	9.3
urban_tv_2018 (en %)	0.1	3.9	8.7	22.2	27.9	100
acces_sante (en %)	48	96.80	100	97.01	100	100

Source : Ined, auteur

Etude des corrélations et statistiques bivariées

La Figure 2 montre les distributions de la variable taux de recours au sein des quintiles des autres variables, ou des catégories pour la variable *accès santé* et *distance aux centres AMP*. Les trois variables pour lesquelles les liens semblent les plus forts au vu de la distribution des taux de recours par quintile sont l'APL pour les médecins gynécologues, le taux d'urbanisation et l'accès santé en moins de 10 minutes. Pour ces variables, le taux de recours est plus élevé pour les quintiles les plus favorisés. Pour l'indice de défavorisation sociale *FDep13*, le cinquième quintile (les territoires de vie les plus défavorisés) est surprenant car il ne suit pas la tendance des quatre premiers quintiles. Pour les quatre

premiers quintiles, la tendance suggère que les territoire de vie socialement favorisés ont des taux de recours plus élevés. Pour la variable de l'APL généraliste, il n'apparaît aucune tendance de distribution des taux de recours entre les différents quintiles. Pour la variable de la distance aux centres AMP, les territoires de vie ayant un centre sur leur territoire (catégorie 1) ont des taux de recours plus important que ceux ne disposant d'un centre qu'au-delà de la deuxième couronne (catégorie 0).

Le Tableau 4 présente les corrélations entre les variables quantitatives du modèle. L'objectif de ce tableau est de montrer les liens existants entre les variables afin de déceler une éventuelle colinéarité. En effet, une forte corrélation n'est pas une condition suffisante pour conclure à la colinéarité mais c'est une condition nécessaire. La corrélation la plus forte est entre la variable d'urbanisation (*urban_tv*) et l'APL pour les médecins gynécologues (*apl_gyn*) avec une valeur de 0,64. La variable la plus corrélée avec le taux de recours est le taux d'urbanisation du TV (0,45). L'accès aux médecins gynécologues se situe en seconde position avec (0,33). À contrario, et cela est plus surprenant, l'accès aux médecins généralistes et le niveau de favorisation sociale ont des corrélations plus faibles avec le taux de recours. Il est intéressant de signaler également que le taux d'urbanisation et l'APL pour les gynécologues ont des corrélations élevées avec presque toutes les autres variables du modèle. Bien que cela ne soit pas suffisant, cela peut alerter sur la présence d'une colinéarité entre ces variables du taux d'urbanisation et de l'APL gynécologue.

Tableau 4 : Table des corrélations entre les variables quantitatives de l'étude

Corrélation	Recours	FDep13	Apl_gyn	Apl_gen	Acces_sante	Urban_tv	Act_amp_tv
Recours	1	-0.11	0.33	0.04	0.28	0.45	0.15
FDep13		1	-0.44	-0.10	-0.20	-0.37	-0.06
Apl_gyn			1	0.47	0.40	0.64	0.08
Apl_gen				1	0.20	0.25	0.04
Acces_sante					1	0.32	0.03
Urban_tv						1	0.06
Act_amp_tv							1

Source : Ined, auteur

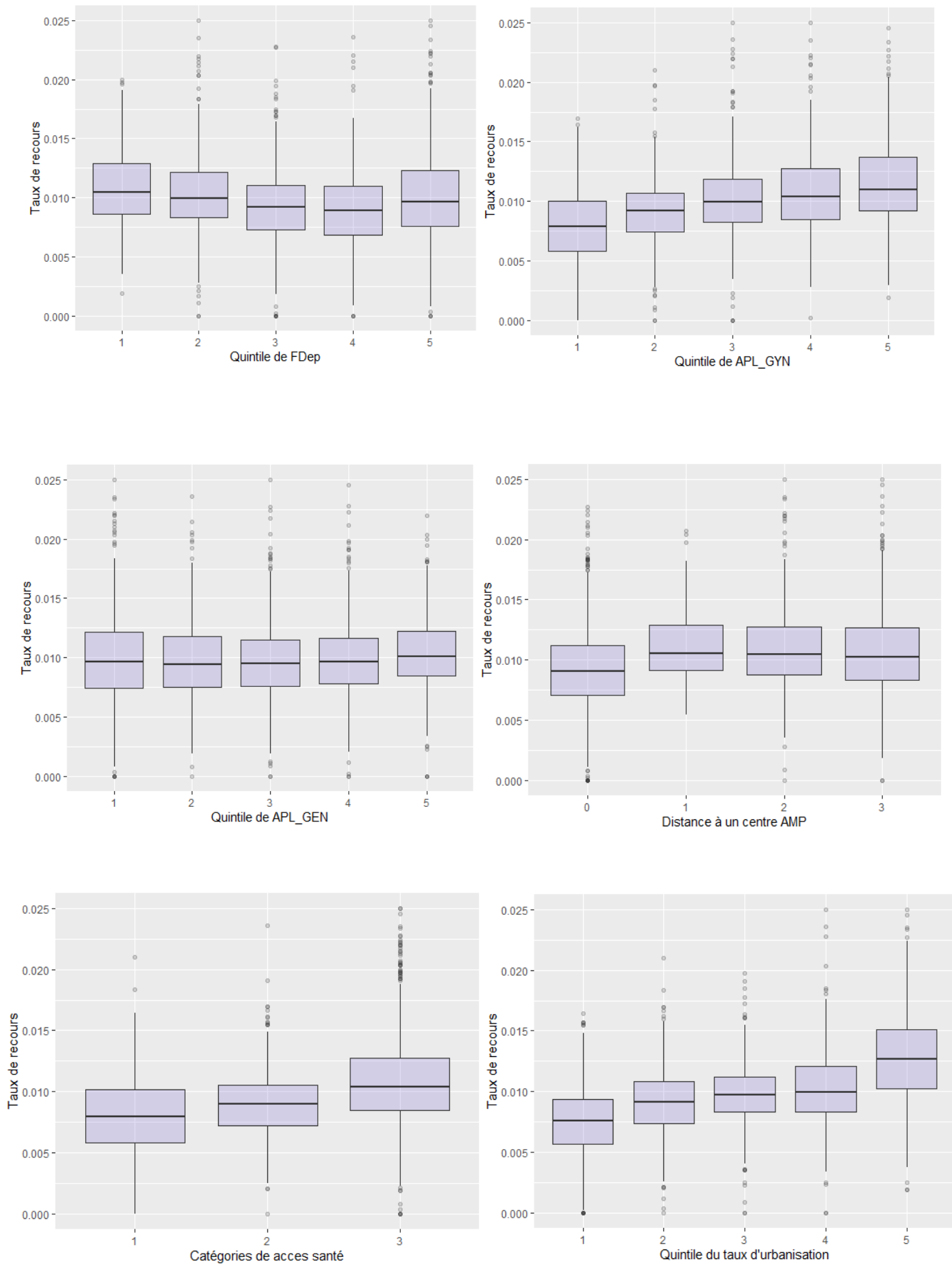


Figure 2 : Distribution des taux de recours en fonction des quintiles (ou catégories) des variables explicatives

Résultats de l'Analyse en Composantes Principales (ACP)

L'objectif de l'ACP est de continuer l'exploration descriptive des variables et de déceler une éventuelle colinéarité entre les variables. La Figure 3 montre la proportion d'inertie portée par chaque axe de l'ACP. Le premier axe se démarque avec une inertie de 37,9 % contre 15,3 % pour le second axe. La proportion d'inertie diminue progressivement par la suite. La Figure 4 représente le graphe des variables. Les deux premiers axes forment le sous-espace sur lequel les variables sont projetées. Dans ce cas, le premier axe est positivement corrélé avec l'indice de défavorisation sociale *fdep13* et négativement corrélé avec le taux d'urbanisation *urban_tv_2018*, l'indice APL gynécologue *apl_gyn* et l'accès à des soins de santé basiques *acces_sante*. Ce premier axe porte une notion de territoires de vie urbains, socialement favorisés, avec des médecins gynécologues, un bon accès global aux soins de santé et un taux de recours supérieur à la moyenne. Le second axe est positivement corrélé avec l'activité des centres AMP et le taux de recours. Cet axe porte donc la notion de territoires de vie avec une forte activité de soins d'infertilité.

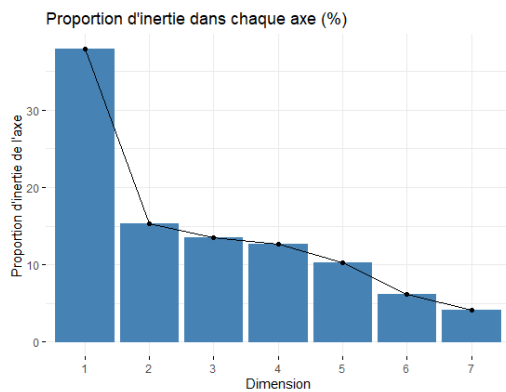


Figure 3 : Proportion d'inertie expliquée par chaque axe de l'ACP

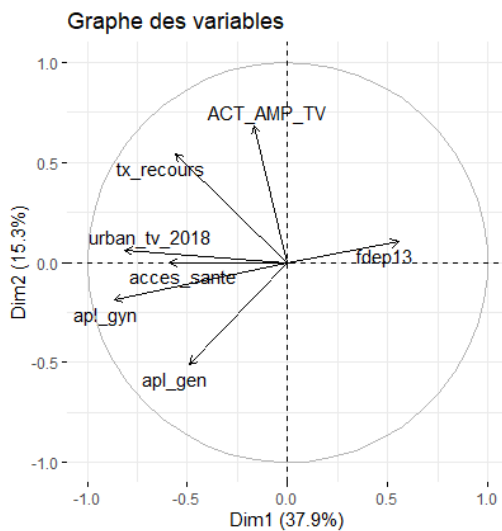


Figure 4 : Graphe des variables de l'ACP

Partie III : Méthodes statistiques

Dès le départ, l'idée de l'étude était de mener une recherche sur le phénomène spatial sous-jacent à l'utilisation des traitements de l'infertilité. Le plan méthodologique est le suivant : un modèle « global » est mis en place sans prendre en compte l'aspect spatial. Puis la structure spatiale est codifiée grâce aux matrices de voisinage. Les statistiques de Moran permettent de tester l'hypothèse d'absence d'autocorrélation spatiale dans les variables et dans les résidus du modèle « global ». Finalement, afin de prendre en compte un possible phénomène d'hétérogénéité spatiale, un modèle RGP est construit. Le langage R est utilisé pour l'analyse statistique avec l'environnement de travail R *Studio*.

3.1 Etude de la colinéarité

En plus de l'étude des corrélations et de l'ACP, une étude de la colinéarité peut être nécessaire à différents moments de la construction méthodologique. Une mesure classique de la multi colinéarité est le facteur d'inflation de la variance (VIF en anglais). En notant vif_j le VIF de la variable j et R_j^2 le carré de la régression de la variable j sur toutes les variables de l'étude sauf elle-même, alors $vif_j = \frac{1}{1-R_j^2}$ (O'Brien 2007). Il n'existe pas dans la littérature un niveau précis à partir duquel les chercheurs considèrent que le VIF est trop élevé. Un seuil commun est de 10 mais il est courant de voir dans la littérature des seuils utilisés de 5 ou de 2,5. Il n'existe pas non plus de test pour la multi colinéarité. Le VIF sera un outil pour détecter la colinéarité et motiver la décision de retirer une variable de l'étude à la lumière d'autres indicateurs allant dans ce sens.

3.2 Modèle de Poisson

Dans cette étude, la variable d'intérêt est une variable du taux de recours. Il est nécessaire de sortir du cadre classique de la régression linéaire simple car l'hypothèse de normalité ne peut pas être vérifiée et de mettre en place des modèles linéaires généralisés. Le premier modèle mis en place fut une régression Beta pour directement modéliser la loi du taux de recours conditionnellement par une loi Beta de paramètres α et β avec la fonction de distribution :

$$f(x; \alpha, \beta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} \mathbb{1}_{(0,1)}(x)$$

Cependant, la régression Beta possède un inconvénient majeur dans cette étude qui est que les valeurs du taux de recours doivent être incluses entre 0 et 1, 0 et 1 exclus de l'intervalle. Or certains territoires de vie ($n=12$ soit 0,004 % des territoires de vie) possèdent des taux de recours nuls ce qui oblige à les exclure de l'étude pour mettre en place cette méthode. Dès lors, il semble préférable d'utiliser une régression de Poisson qui permet de prendre en compte ces territoires où aucune femme n'a eu recours à ces traitements. Ce modèle est néanmoins pertinent uniquement si le nombre de territoires avec un taux nul est faible, sinon un problème connu sous le nom du « zero inflation » peut apparaître. Dans notre cas, ce nombre étant faible (12 soit 0,4 % des territoires de vie), un modèle de Poisson classique peut être mis en place.

Un modèle de poisson modélise le paramètre de la loi de Poisson par la relation log-linéaire suivante :

$$\ln \lambda(\mathbf{X}) = \beta \cdot \mathbf{X}$$

Une régression de Poisson s'appuie sur une variable de comptage. Dans notre étude la variable de comptage est le nombre d'acheteuses de traitements de l'infertilité `eff_stim`. Ce nombre est divisé par le nombre de femmes âgées de 18 à 49 ans dans le territoire de vie `eff_tot` pour obtenir la variable taux utilisée dans la régression beta `tx_recours`. La fonction de lien utilisée est une fonction de lien log

et l'écriture du modèle est présentée ci-après. β représente le vecteur des coefficients associés aux régresseurs et X la matrice des valeurs de ces régresseurs.

$$\log\left(\frac{\text{eff_stim}}{\text{eff_tot}}\right) = \beta \cdot X$$

$$\log(\text{eff_stim}) - \log(\text{eff_tot}) = \beta \cdot X$$

$$\log(\text{eff_stim}) = \log(\text{eff_tot}) + \beta \cdot X$$

Dans ce modèle, la variable $\log(\text{eff_tot})$ est appelé un **offset** du modèle. Le coefficient associé à cet offset est égal à 1. Une hypothèse de ce modèle est que la loi du nombre d'acheteuses de traitements sachant X suit une loi de poisson. Or si une variable aléatoire Y suit une loi de Poisson, son espérance est égale à sa variance : $E[Y|X] = \text{Var}(Y|X)$. En pratique, il peut y avoir une sur-dispersion, c'est-à-dire $E[Y|X] < \text{Var}(Y|X)$ ou une sous-dispersion $E[Y|X] > \text{Var}(Y|X)$ qu'il faut corriger.

Ecriture du modèle

Une limite de ce modèle est que les résultats avec les variables quantitatives sont difficilement interprétables. En effet, une variable comme la défavorisation sociale n'a pas d'unité à proprement parler. Il est plus pertinent de transformer ces variables en variables catégorielles. Pour cela, la variable est découpée en quintile et chaque territoire de vie est assimilé à un de ces quintiles.

Dans la pratique, le logiciel *R* crée des variables binaires lors de l'implémentation d'un modèle linéaire généralisé. Notons $1_{(fdep13=1)}$ l'indicatrice d'appartenance au premier quintile de la variable De défavorisation sociale. L'équation du modèle est la suivante :

$$\begin{aligned} \log(\text{eff}_{stim}) = \log(\text{eff}_{tot}) &+ \sum_{i=1}^5 \alpha_i 1_{(fdep13=i)} + \sum_{i=1}^5 \beta_i 1_{(apl_{gyn}=i)} + \sum_{i=1}^5 \gamma_i 1_{(apl_{gen}=i)} \\ &+ \sum_{i=1}^5 \delta_i 1_{(urban=i)} + \sum_{i=0}^3 \theta_i 1_{(dist_{amp}=i)} + \sum_{i=1}^3 \mu_i 1_{(acce_{sante}=i)} + \rho \text{ACTAMP} \end{aligned}$$

Avec $\alpha_1 = \beta_1 = \gamma_1 = \delta_1 = \theta_0 = \mu_1 = 0$ pour s'assurer que la matrice $X'X$ est de plein rang, X étant la matrice des régresseurs.

Interprétation des coefficients

Pour avoir des résultats directement interprétables, l'influence des régresseurs est mesurée à partir du **rate-ratio**. Pour deux individus présentant les caractéristiques x_1 et x_2 le rate-ratio se définit comme :

$$\text{RR}(x_1, x_2) = \frac{\lambda(x_1)}{\lambda(x_2)} = \frac{\exp(x_1 \cdot \beta)}{\exp(x_2 \cdot \beta)} = \exp(\beta \cdot (x_1 - x_2))$$

Notamment, dans le cas de deux individus qui ne diffèrent que par la valeur du j^{eme} régresseur, le rate-ratio devient :

$$\text{RR}(x_1, x_2) = \exp(\beta_j \cdot (x_{1j} - x_{2j}))$$

Et si ce j^{eme} régresseur est un régresseur binaire, le rate-ratio s'écrit

$$\text{RR}(1,0) = \exp(\beta_j)$$

Et en notant $IC(\hat{\beta}_j) = [g, d]$ l'intervalle de confiance de niveau $1 - \alpha$ de l'estimateur du coefficient associé à ce régresseur, l'intervalle de confiance du rate-ratio est $IC(RR(1,0)) = [\exp(g), \exp(d)]$.

3.3 Matrice de distances

Afin d'étudier l'effet spatial dans l'étude, il est nécessaire de construire une matrice de pondérations qui s'appuie elle-même sur une notion de distance et de voisinage. Il existe une multiplicité de matrices de pondérations possibles avec les mêmes données, cependant la méthode s'organise autour des mêmes étapes. La première étape est de définir la notion de voisins, puis de construire une matrice de voisinage et finalement, à partir de la matrice de voisinage, de construire la matrice de pondération et la normaliser.

Matrice basée sur la contiguïté

La première matrice de pondération utilisée se base sur la notion de contiguïté. La contiguïté peut prendre deux formes.

- La première est la contiguïté « Rook » : deux zones sont voisines si elles partagent un segment en commun.
- La seconde est la contiguïté « Queen » : deux zones sont voisines dès lors qu'elles se touchent même ponctuellement.

La matrice de pondération, notée P , est de taille $n \times n$ où n est le nombre de territoires de vie. Le poids $p_{i,j}$ entre l'individu i et l'individu j est défini comme $p_{\{i,j\}} = 1$ si i et j sont contigus et $p_{\{i,j\}} = 0$ sinon. Enfin, cette matrice est normalisée selon la méthode "W" qui désigne une normalisation en ligne. Chaque poids de l'individu i est divisé par le nombre de voisins de cet individu, de tel sorte que $\sum_{j=1}^n p_{\{i,j\}} = 1$. Cette normalisation est populaire car elle offre la possibilité d'interpréter la quantité $\sum_{j=1}^n p_{\{i,j\}} \cdot x_j$ comme la moyenne de la variable x pour les voisins de l'individu i . En revanche, certains individus peuvent être plus influents que d'autres.

Matrice basée sur la distance entre deux points

Deux autres matrices de pondération sont utilisées dans l'étude et elles reposent sur la distance entre les points dans le plan. Concrètement, les points sont les barycentres des territoires de vie. Par la suite, la distance entre deux points i et j est notée $d(i,j)$. Soit n le nombre total de territoires de vie. La première pondération basée sur la distance utilise la distance des voisins relatifs. Les notations sont inspirées du Chapitre 2 du Manuel d'analyse spatiale de l'Insee (Bellefon, Loonis et Le Gleut 2018). Dans cette distance, deux points i et j sont voisins si et seulement si :

$$\forall k \in \llbracket 1, n \rrbracket, \quad k \neq \{i, j\},$$

$$d(i, j) \leq \max\{d(i, k), d(j, k)\}$$

L'intuition géométrique de cette distance est que deux points i et j sont voisins si l'intersection des deux cercles centrés en i et j de rayon $d(i,j)$ ne contient pas d'autre points que i et j . Cette vision géométrique est présentée par Fergani et Soukhal dans un article présentant un algorithme incrémental des graphes de voisins relatifs (Fergani et Soukhal 2015).

La seconde pondération issue de la distance entre les points est celle des k plus proches voisins. La principale difficulté de cette distance est de choisir la valeur du k qui « reflète les vraies relations

spatiales sous-jacente » (Bellefon, Loonis et Le Gleut 2018). Ces deux distances (voisins relatifs et k plus proches voisins) ont l'intérêt de ne pas laisser de TV sans voisins.

3.4 Tester la présence d'autocorrélation spatiale

Nature des effets spatiaux

Les effets spatiaux peuvent prendre plusieurs formes. D'une part, il existe les notions d'autocorrélation spatiale et d'hétérogénéité spatiale.

- **L'autocorrélation spatiale** désigne le fait que les individus proches spatialement ne sont pas indépendants, ce qui viole une hypothèse des modèles « globaux » de régression de Poisson et les rend non-pertinents. La présence d'autocorrélation spatiale rend également invalide les tests d'hétéroscédasticité (Le Gallo 2004). Une autocorrélation spatiale peut émerger pour plusieurs raisons. D'une part les données peuvent être « affectées par des processus qui relient des lieux différents » et d'autre part « elle peut provenir d'une mauvaise spécification du modèle » (Le Gallo 2004).
- **La notion d'hétérogénéité spatiale** est différente, elle repose sur l'idée que la nature du modèle diffère d'une zone géographique à une autre. Autrement dit, les variables explicatives peuvent être les mêmes sur tout le territoire, mais avoir une influence différente d'une région à une autre. Les variances peuvent également être différentes d'une zone géographique à une autre, auquel cas il est question d'hétéroscédasticité spatiale.

Les relations qui lient autocorrélation spatiale (par processus de génération de données spatiales ou par mauvaise spécification) et l'hétérogénéité spatiale (instabilité spatiale des coefficients ou des variances) sont complexes. Ces phénomènes peuvent parfois être concomitants. De plus, l'hétérogénéité spatiale peut entraîner l'émergence d'une autocorrélation spatiale : « si les relations variables dans l'espace sont modélisées à l'aide d'une régression « globale », les termes d'erreurs peuvent être spatialement autocorrélés » (Le Gallo 2004).

Concrètement, la stratégie mise en place dans cette étude pour analyser le phénomène spatial est la suivante : premièrement, tester la présence d'autocorrélation spatiale dans les variables de l'étude et dans les termes d'erreur du modèle de Poisson « global ». Deuxièmement, mettre en place un modèle de régression géographiquement pondéré (RGP) pour étudier l'hétérogénéité spatiale et plus particulièrement l'instabilité spatiale des coefficients. Cela reviendra à tester la non-stationnarité spatiale des coefficients et donc du phénomène spatial.

Présence de l'autocorrélation spatiale

Pour débiter, il faut s'assurer de la présence d'autocorrélation spatiale dans les modèles. Pour cela, plusieurs éléments sont à dispositions. Tout d'abord, les cartes présentées auparavant (ex : Figure 1) montrent visuellement des disparités sur le territoire français métropolitain et la présence de zones dites chaudes (où les valeurs sont élevées et entourées de valeurs élevées) et de zones froides (valeurs faibles entourées de valeurs faibles). Pour aller au-delà de l'aspect visuel, il existe des indices et tests statistiques dont le plus connu est celui du I de Moran. Ces indices servent à mesurer la ressemblance globale des valeurs proches. Pour les calculer, il est donc nécessaire de définir une notion de proximité à l'aide de matrices de pondérations de la distance.

Une fois la matrice de pondération construite, il est possible de calculer le I de Moran. L'indice du I de Moran sert à établir la présence d'autocorrélation spatiale. Il est défini comme suit :

$$I = \frac{N}{\sum_i \sum_j p_{i,j}} \frac{\sum_i \sum_j p_{i,j} (y_i - \hat{y})(y_j - \hat{y})}{\sum_i (y_i - \hat{y})^2}$$

Dans cette équation $p_{i,j}$ est la pondération entre l'individu i et l'individu j , \hat{y} est la moyenne de la variable y . De cet I de Moran, il est possible de calculer la statistique de test suivante :

$$\frac{I - E[I]}{\sqrt{Var(I)}}$$

Sous l'hypothèse nulle de l'absence d'autocorrélation spatiale, cette statistique de test suit une loi normale centrée-réduite.

I de Moran local

Derrière le calcul du I de Moran, il y a l'hypothèse que le phénomène d'autocorrélation spatiale est homogène sur le territoire. Si cela n'est pas vrai, il est intéressant de calculer des indices locaux d'autocorrélation spatiale pour mieux se rendre compte du phénomène sur le territoire d'étude. Ces indicateurs sont appelés *Local Indicators of Spatial Association*, notés LISA. Un indicateur LISA populaire est celui de I de Moran local, noté I_i dont l'équation est la suivante.

$$I_i = (y_i - \hat{y}) \sum_j p_{i,j} (y_j - \hat{y})$$

Pour rappel, avec la normalisation "W" de la matrice de pondération, la quantité $\sum_{j=1}^n p_{i,j} y_j$ s'interprète comme la moyenne sur la variable y des voisins de l'individu i . Il vient naturellement que si $I_i > 0$ alors l'individu i et ses voisins prennent des valeurs similaires sur y , soit plus faibles que la moyenne, soit plus élevés. A contrario, si $I_i < 0$ alors les valeurs de l'individu i et de ses voisins diffèrent.

3.5 Modèle de régression géographiquement pondérée

Par ailleurs, le phénomène spatial peut prendre la forme d'une hétérogénéité spatiale. L'hypothèse de l'hétérogénéité spatiale consiste à dire que les coefficients des facteurs explicatifs varient dans l'espace, auquel cas un modèle de régression « simple » ne suffit pas à capter ces variations spatiales. En effet, « les paramètres estimés (des régressions simples) sont supposés constants dans l'espace ce qui peut être une hypothèse très discutable dans de nombreuses circonstances » (Brunsdon, Fotheringham et Charlton 1996)¹⁷. Dans l'article de Harris et ses collègues (Harris, et al. 2016), l'idée que les coefficients puissent évoluer dans l'espace est pris en compte par des indicatrices de zones géographiques. Cette méthode est couramment utilisée dans la littérature mais dans cette étude la non-stationnarité spatiale est étudiée à l'aide d'une régression géographiquement pondérée (RGP). L'idée de la RGP est de réaliser un modèle de Poisson par territoire de vie (2719 au total), « chaque observation étant pondérée par une fonction décroissante de sa distance au point d'estimation » (Bellefon et Floch, Chapitre 9 - Régression géographiquement pondérée 2018). Plus un territoire de vie sera éloigné du territoire où le modèle est estimé, moins il aura de poids dans l'estimation. Soit (u_i, v_i) les coordonnées du territoire de vie i , alors il est nécessaire de construire une matrice de pondération pour ce TV : $W_{(u_i, v_i)}$. Cette matrice dépend d'une fonction de noyau, et plus précisément des paramètres suivants :

- L'adaptativité du noyau
- La forme du noyau
- La taille de la bande-passante

¹⁷ Traduit par l'auteur, à l'origine: « The resulting statistics or parameter estimates are assumed to be constant across space although this might be a very questionable assumption to make in many circumstances »

L'adaptativité du noyau

Un noyau adaptatif ne sera pas le même pour tous les territoires de vie, à contrario du noyau fixe. La nécessité d'un noyau adaptatif apparaît surtout lorsque la taille des territoires varie et, de facto, leurs éloignements. Opter pour un noyau fixe exposerait au risque d'avoir un nombre d'individus dans chaque estimation très différent. **Il semble préférable de choisir un noyau adaptatif dans cette étude.**

La forme du noyau

Il existe deux familles de noyaux : ceux qui donnent un poids non nul à chaque individu et ceux pour lesquels les individus ont un poids nul au-delà d'une certaine distance (ils ne rentrent en plus en compte dans l'estimation du modèle). La première famille est appelée « noyaux continus » et la seconde « noyaux à support compact ». Dans cette étude, **le noyau de la seconde famille « bi-square » est utilisé.** Il est recommandé par Marie-Pierre de Bellefon et Jean-Michel Floch dans le Manuel d'analyse spatiale car c'est le noyau qui optimise le temps de calcul (Bellefon et Floch, Chapitre 9 - Régression géographiquement pondérée 2018). La forme du noyau bi-square s'appuie sur la distance entre le territoire de vie i et le territoire de vie j notée $d_{i,j}$. Elle se formule comme suit, h est la notation pour la bande-passante qui fait l'objet du paragraphe suivant :

$$w(d_{i,j}) = \begin{cases} \left(1 - \left(\frac{d_{i,j}}{h}\right)^2\right)^2 & \text{si } d_{i,j} < h \\ 0 & \text{sinon} \end{cases}$$

La bande-passante : h

La bande passante est un paramètre de la régression RGP qui intervient directement dans le calcul de la matrice de pondération précédente. Notamment, au-delà de cette bande-passante les individus ont des poids nuls et ne sont pas pris en compte dans l'estimation des coefficients. La bande-passante est « le paramètre dont le choix a la plus forte influence sur les résultats » (Bellefon et Floch, Chapitre 9 - Régression géographiquement pondérée 2018). Le choix de la bande-passante peut se faire à l'aide de deux critères à minimiser : le critère de validation croisée et le critère de l'AIC corrigé. Le critère de validation croisée maximise la capacité prédictive du modèle quand le critère de l'AIC corrigé donne un compromis entre complexité et prédiction à l'aide d'un terme de pénalité (Nakaya, et al. 2005).

La valeur de la bande-passante peut donner une indication sur la structure du phénomène spatial. En effet, une bande-passante optimale (au sens d'un des deux critères) qui prend tout le territoire signifie sans doute qu'il n'est pas nécessaire de faire une RGP. A l'inverse, une valeur très petite de la bande-passante ne permet pas d'obtenir des estimations fiables car le nombre d'individus inclus dans les régressions sont trop faibles (Nakaya, et al. 2005).

Estimation par maximum de vraisemblance

Finalement, l'écriture du modèle est la suivante.

$$\log(\text{eff_stim}) = \log(\text{eff_tot}) + \beta_{(u_i, v_i)} \cdot X$$

L'estimation des coefficients $\beta_{(u_i, v_i)}$ se fait par maximum de vraisemblance géographiquement pondérée. Dans la formule suivante et pour rendre l'écriture plus lisible, le terme O_j désigne en réalité l'effectif d'acheteuses de traitements dans le TV j qui était auparavant noté eff_stim . Le terme $\hat{O}_j(\beta_{(u_i, v_i)})$ désigne lui le nombre d'acheteuses prédit dans le territoire de vie j avec les paramètres de la régression au point i (Nakaya, et al. 2005).

$$\max(L_{u_i, v_i}) = \sum_j^N (-\hat{O}_j(\beta_{(u_i, v_i)}) + O_j \log \hat{O}_j(\beta_{(u_i, v_i)})) w(d_{i,j})$$

Il faut remarquer que la vraisemblance est « géographiquement pondérée, les poids étant une fonction de Kernel centrée en (u_i, v_i) »¹⁸ (Nakaya, et al. 2005).

Interprétation des coefficients de la RGP

La RGP consiste à réaliser une régression de Poisson pour chaque TV avec le paramétrage adéquat. Dans cette étude, elle sera réalisée à l'aide du package *GWmodel*¹⁹ et des fonctions *gwr.generalised* ou *gwr.basic* qui fournissent exactement les mêmes résultats. Il y a donc 2719 équations de régression avec 12 coefficients estimés et les données sont stockées dans un objet de type *Spatial Polygon Data Frame*. Chaque coefficient s'interprète comme celui d'une régression de Poisson « classique ». L'intérêt de la méthode est d'observer, ou non, la variabilité dans l'espace des coefficients.

Cela permet notamment de présenter les résultats sous forme de cartes, ce qui est intéressant puisque que la méthode RGP est une méthode où l'aspect géographique est important. Avant de regarder les cartes, il est possible de regarder la distribution des coefficients à l'aide des quartiles et de la moyenne par exemple. Je propose également de réaliser une ACP sur ces données afin de visualiser différemment les résultats et d'identifier les variables explicatives ayant des impacts remarquables (dans le sens plus ou moins fort que la moyenne) conjointement. Les cartes permettent ensuite de visualiser cette variabilité et d'identifier des zones géographiques où l'impact d'une variable est plus faible ou plus fort.

¹⁸ Traduit par l'auteur, texte original: « likelihood is geographically weighted, with the weights being a kernel function centred on u_i . »

¹⁹ Voir Annexe pour l'ensemble des packages utilisés lors de l'étude

Partie IV : Résultats des modèles multivariés

Les premiers modèles de régressions

Le Tableau 5 présente les résultats du modèle de Poisson et du modèle de Poisson corrigé de la sur-dispersion. Pour chaque groupe de quintiles, la valeur du VIF est mentionnée. Le VIF pour les quintiles d'urbanisation est de 4,72. Au regard du premier axe de l'ACP qui porte la notion conjointe d'urbanisation et de favorisation sociale et du VIF élevé de la variable du taux d'urbanisation, il semble qu'un problème de multicolinéarité soit possible entre les variables taux d'urbanisation, la défavorisation sociale et APL pour les gynécologues. Si ces variables mesurent en effet un même phénomène, il semble que ce soit plutôt la défavorisation sociale du territoire de vie et l'accès aux médecins gynécologues qui puissent avoir une influence sur le comportement des femmes vis à vis de leurs environnements. Il a été décidé de retirer la variable du taux d'urbanisation de l'analyse pour se concentrer sur la défavorisation sociale et l'accès aux médecins gynécologues.

Tableau 5 : Résultats des modèles complets de Poisson et Poisson corrigé de la sur-dispersion

Variable	Quintile/catégorie	Poisson	Poisson corrigé	VIF
FDep13	2/1	0.079 (***)	0.079 (***)	1.43
	3/1	0.070 (***)	0.070 (***)	
	4/1	0.090 (***)	0.090 (***)	
	5/1	0.199 (***)	0.199 (***)	
APL généraliste	2/1	-0.062 (***)	-0.062 (***)	1.50
	3/1	-0.075 (***)	-0.075 (***)	
	4/1	-0.137 (***)	-0.137 (***)	
	5/1	-0.241 (***)	-0.241 (***)	
APL gynécologue	2/1	0.059 (***)	0.059 (*)	3.55
	3/1	0.169 (***)	0.169 (***)	
	4/1	0.158 (***)	0.158 (***)	
	5/1	0.187 (***)	0.187 (***)	
Urbanisation	2/1	0.092 (***)	0.092 (***)	4.72
	3/1	0.187 (***)	0.187 (***)	
	4/1	0.246 (***)	0.246 (***)	
	5/1	0.445 (***)	0.445 (***)	
Distance AMP	1/0	-0.075 (***)	-0.075	2.53
	2/0	0.025 (**)	0.025 (***)	
	3/0	0.075 (***)	0.075 (***)	
Accès santé	2/1	0.107 (***)	0.107 (**)	2.13
	3/1	0.118 (***)	0.118 (***)	
Activité AMP	-	0 (***)	0 (*)	1.44
Constante	-	-5.032	-5.032	-
Log vraisemblance	-	-12166	-	-

Code significativité : 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0.1

Source : Ined, auteur

Autocorrélation spatiale : statistique de Moran

L'objectif de cette sous-partie est de présenter les résultats liés à l'autocorrélation spatiale. Le Tableau 6 présente les statistiques du test de Moran sur les variables quantitatives de l'étude ainsi que sur les résidus des modèles de Poisson réalisés. **L'hypothèse nulle d'absence d'autocorrélation spatiale est rejetée pour toutes les variables.** Cela confirme le constat visuel lors de l'exploration des données mais n'est pas suffisant pour affirmer la pertinence des modèles spatiaux. En effet, toute l'autocorrélation spatiale des taux de recours pourrait être expliquée par celles des variables explicatives. Autrement dit, le modèle de Poisson « simple » pourrait expliquer tout l'aspect spatial du phénomène. Si tel était le cas, alors les résidus des régressions de Poisson ne seraient pas autocorrélés spatialement. Or le Tableau 6 montre également que **l'hypothèse nulle est rejetée pour les résidus des modèles de Poisson complet et avec sélection de variables.** Il est donc nécessaire de réaliser des modèles spatiaux pour prendre en compte cette dépendance spatiale. La Figure 5 représente le graphe de Moran pour la variable des taux de recours. Cette figure permet également de se rendre compte que le phénomène n'est pas aléatoire, il y a présence d'une autocorrélation spatiale positive. Cela se remarque par la régression linéaire du nuage de points à la pente positive. La Figure 6 représente la carte des I de Moran locaux pour les TV en France métropolitaine. Les valeurs élevées ou faibles des statistiques sont respectivement dans une couleur sombre ou claire. Il existe des zones de regroupement de valeurs élevées (région Ile-De-France, Champagne-Ardenne), ainsi que des zones de regroupements de valeurs faibles (sud-ouest). Cette carte doit être lue avec la carte présentée Figure 7 qui correspond aux p-valeurs associées à ces I de Moran locaux. Les régions citées précédemment ressortent comme étant significatives, cela confirme la présence de « clusters » significatifs.

Tous ces résultats sont autant d'indications qu'il existe une autocorrélation sur le territoire à prendre en compte dans les modèles. Dans la partie suivante, l'accent est mis sur l'interprétation du modèle RGP qui explore l'hypothèse d'une hétérogénéité spatiale.

Tableau 6 : Récapitulatif des tests de Moran sur les variables et résidus des modèles

	Statistique du test	P-valeur associée
Taux de recours	37,28	< 0,001
APL gynécologue	55,06	< 0,001
APL généraliste	37,64	< 0,001
FDep13	57,57	< 0,001
Taux d'urbanisation	68,32	< 0,001
Modèle complet	36,38	< 0,001
Sélection de variables	36,38	< 0,001

Note de lecture : La statistique de test pour le test de Moran est de 37,28 pour la variable du taux de recours. La p-valeur associée à ce même test est inférieure à 0,1 %. Il faut rejeter l'hypothèse nulle d'autocorrélation spatiale.

Source : Ined, auteur

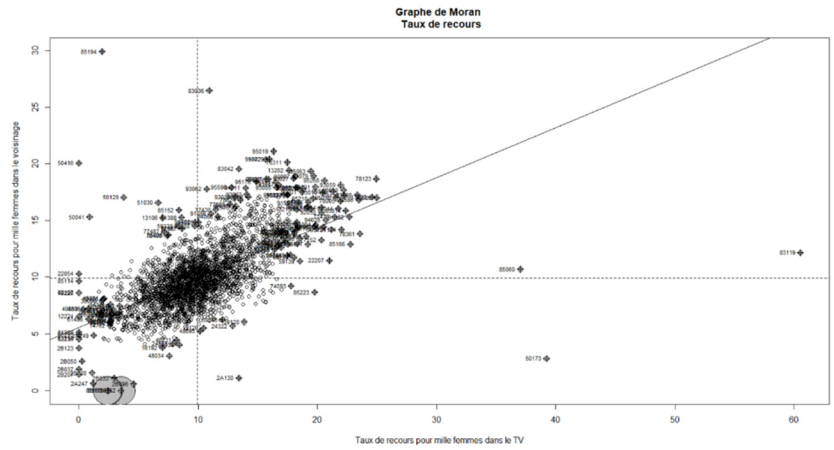


Figure 5 : Graphe de Moran du taux de recours pour mille femmes

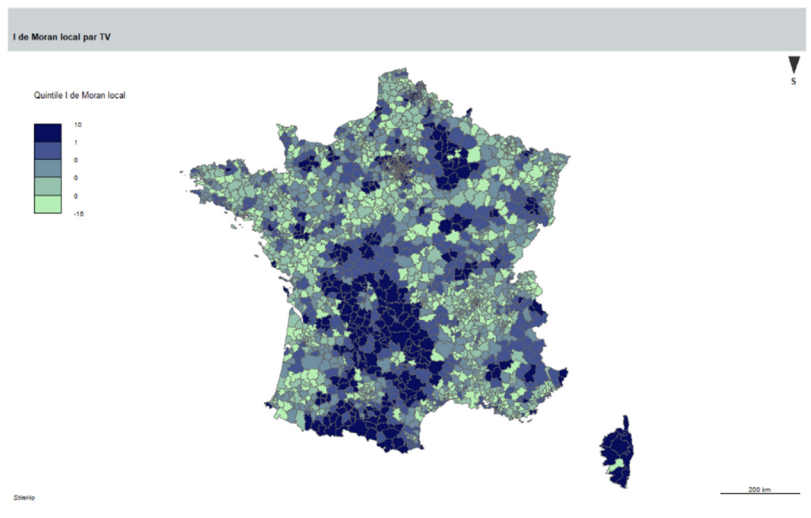


Figure 6 : I de Moran local par territoire de vie en France métropolitaine

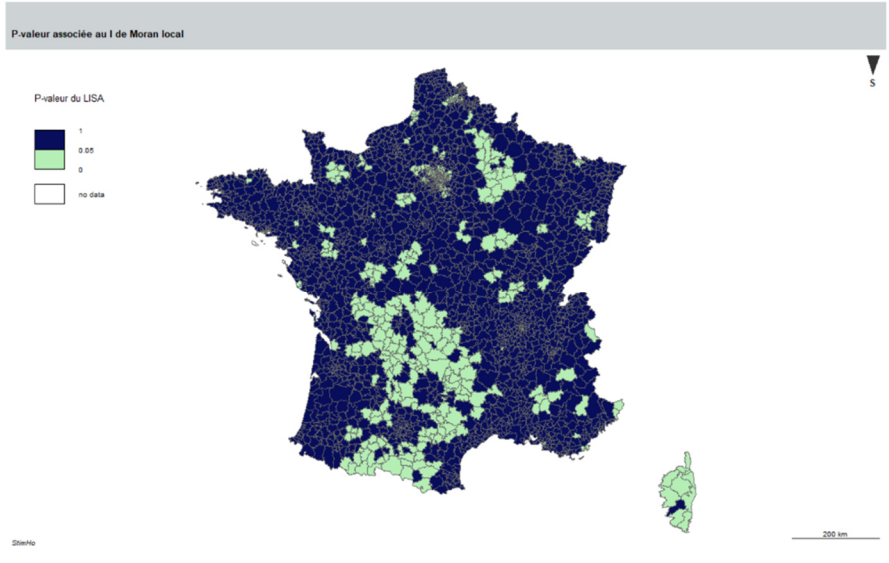


Figure 7 : P-valeurs associées aux I de Moran locaux

Modèle de régression géographiquement pondérée

Pour une présentation plus lisible des résultats de la régression géographiquement pondérée (RGP), les valeurs des coefficients seront représentées par des cartes. Pour réduire le nombre de cartes, la décision a été prise de redécouper les variables de l'étude en terciles. Le tercile de référence pour l'interprétation des résultats est le premier tercile. Le Tableau 7 permet de comparer les résultats de la régression de Poisson globale avec les résultats de la RGP. Pour ce faire, il présente quelques statistiques descriptives des coefficients obtenus par TV. A cela s'ajoute les résultats de l'ACP des résultats de la RGP avec la proportion d'inertie par axe (Figure 8) et le graphe des variables (Figure 9).

Tableau 7 : Comparaison entre le modèle de Poisson global et la RGP

Modèle		Poisson	RGP					
Variables		Coefficients	Min	1 ^{er} quar	Médiane	Moyenne	3 ^e quar	Max
FDep13	2/1	0,02	-0,06	0,11	0,22	0,23	0,32	0,57
	3/1	0,12 (***)	-0,10	0,06	0,13	0,20	0,29	0,72
APL_GYN	2/1	0,19 (***)	0,36	0,51	0,58	0,59	0,69	0,84
	3/1	0,29 (***)	0,59	0,90	1,00	0,99	1,09	1,37
APL_GEN	2/1	-0,06 (***)	-0,06	0,07	0,12	0,12	0,18	0,21
	3/1	-0,18 (***)	-0,39	-0,03	0,14	0,09	0,22	0,34
DIST_AMP	1/0	0,02	1,03	1,29	1,39	1,51	1,77	2,26
	2/0	0,09 (***)	-0,08	0,11	0,38	0,32	0,48	0,65
	3/0	0,10 (***)	-0,13	0,01	0,09	0,08	0,15	0,27
Accès santé	2/1	0,17 (***)	0,07	0,38	0,46	0,46	0,55	0,73
	3/1	0,31 (***)	-0,08	0,31	0,46	0,46	0,62	0,84

Code significativité : 0 '***' 0,001 '**' 0,01 '*' 0,05 '.' 0,1

Note de lecture : Le coefficient associé au troisième tercile par rapport au premier dans la régression de Poisson global corrigée de la sur-dispersion est de 0,29. Il est significatif au seuil de 0,1 %. La médiane des coefficients pour les 2719 régressions géographiquement pondérées pour ce même coefficient est de 1,00.

Source : Ined, auteur

Figure 8 : Proportion d'inertie dans chaque axe de l'ACP (RGP)

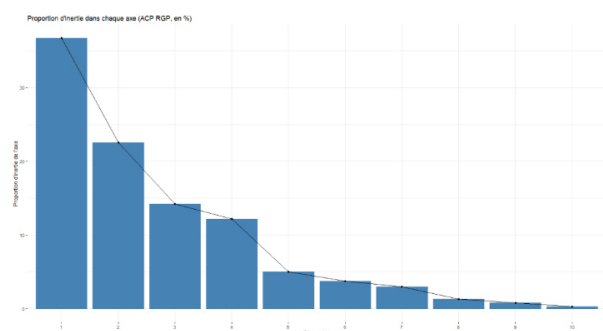
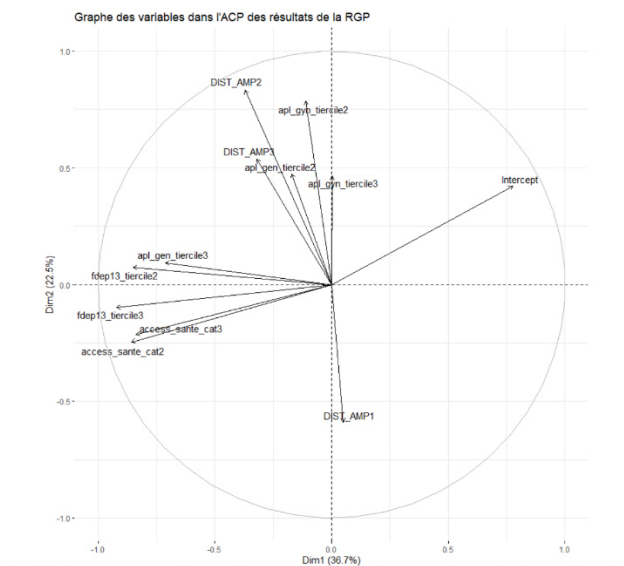


Figure 9 : Graphe des variables des résultats de la RGP



L'accessibilité aux médecins gynécologues présente l'impact positif le plus fort

La régression géographiquement pondérée (RGP) montre que la prise en compte de l'hétérogénéité spatiale est nécessaire pour comprendre le phénomène de recours aux traitements de l'infertilité. En effet, les cartes Figure 15, Figure 16 (en annexes, suivre lien) présentent une variabilité spatiale des coefficients de la régression. Autrement dit, l'impact des variables explicatives varie à travers le territoire ce qui traduit une instabilité spatiale. Le Tableau 7 permet de comparer les résultats de la RGP avec les résultats du modèle de Poisson équivalent. Le premier résultat marquant est que pour certaines variables, les estimations peuvent être négatives ou positives. C'est le cas pour au moins un tercile de toutes les variables sauf l'APL aux médecins gynécologues. L'impact de l'APL gynécologue semble le plus stable à travers le territoire, les minimums étant assez éloignés de zéro (0,36 pour le deuxième tercile par rapport au premier et 0,59 pour le troisième tercile par rapport au premier). Cet impact est positif sur le taux de recours aux traitements de l'infertilité sur tout le territoire. Il est assez élevé avec une moyenne de 0,59 pour le deuxième tercile et de 0,99 pour le troisième. **Toutes choses égales par ailleurs, une bonne accessibilité aux médecins gynécologues favorise le recours aux traitements de l'infertilité dans tous les territoires de vie en France métropolitaine.**

Un impact de l'accessibilité aux médecins généralistes plus équivoque

L'impact moyen de l'APL généraliste sur le territoire est positif (0,12 et 0,09 respectivement pour le deuxième tercile et le troisième tercile par rapport au premier), mais les coefficients varient sur le territoire de France métropolitaine entre le négatif et le positif. Les minimums des deuxième et troisième terciles sont respectivement de -0,06 et -0,39 au premier tercile (Tableau 7). La Figure 17 et la Figure 18 sont les cartes des coefficients associés aux deuxième et troisième terciles de l'APL généraliste sur le territoire de France métropolitaine. Les valeurs les plus élevées se trouvent au sud de la France, en région parisienne et en Alsace. Ces zones géographiques n'ont pas la même accessibilité aux médecins généralistes. Alors que l'Alsace et le sud de la France ont plutôt des valeurs élevées pour l'APL généraliste, ce n'est pas le cas pour la région parisienne. L'effet de l'APL généraliste est donc plus équivoque sur le territoire. Cela n'est pas un résultat surprenant au vu de l'étude des corrélations et des statistiques bivariées une corrélation faible entre l'APL généraliste et le taux de recours aux traitements de l'infertilité.

Les catégories 2 et 3 de la variable « accès santé » ont des résultats assez similaires : par rapport à la catégorie 1, ils favorisent l'accès aux traitements de l'infertilité. Pour rappel, la catégorie 1 rassemble les territoires de vie où moins de 90 % de la population a un accès en moins de dix minutes à une pharmacie et à un médecin généraliste. La catégorie 2 rassemble les territoires de vie ayant cette proportion comprise entre 90 et 99 %. La catégorie 3 sont les territoires de vie à 100 %. Ce résultat montre qu'avoir plus de 10 % de la population d'un territoire de vie à plus de 10 minutes d'une pharmacie et d'un médecin généraliste défavorise l'accès aux traitements de l'infertilité.

La distance aux centres AMP joue un rôle

La RGP permet d'obtenir des résultats vis-à-vis de la distance à un centre AMP. Pour rappel, la première catégorie (catégorie 1) regroupe les territoires de vie ayant un centre AMP en leur sein ; la deuxième catégorie (catégorie 2) regroupe les territoires de vie dont le centre AMP le plus proche est situé en première couronne de ces territoires de vie ; la troisième catégorie (catégorie 3) regroupe ceux dont le centre AMP le plus proche est situé en deuxième couronne ; la dernière catégorie (catégorie 0) regroupent les territoires de vie restants (ceux dont le centre d'AMP le plus proche est situé au-delà de la deuxième couronne). La moyenne des coefficients est la plus élevée pour la catégorie 1/0 (1,51). La moyenne des coefficients associés à être en première couronne d'un centre AMP (2/0) est de 0,32 et celle d'être en deuxième couronne (3/0) est de 0,08. Cela est un faisceau convergent d'indices pour conclure que, toutes choses égales par ailleurs, **plus la distance au centre AMP le plus proche diminue, plus cela favorise le recours aux traitements de l'infertilité.**

Le résultat inattendu de la défavorisation sociale

L'intérêt de la RGP est, pour finir, de mieux comprendre l'impact de la défavorisation sociale dans le recours aux traitements de l'infertilité. La régression de Poisson global estime les coefficients du deuxième et troisième tercile respectivement à 0,02 (sans être significatif à un seuil de 10 %) et 0,12. Ce résultat est inattendu car il signifie que vivre dans un TV socialement défavorisé impact positivement le recours aux soins. C'est le contraire du résultat de Harris et ses collègues (Harris, et al. 2016) bien que le résultat de Harris et al concerne davantage l'accès aux traitements de deuxième intention (IA, FIV). Les cartes en Figure 24 et Figure 25 montrent la répartition spatiale des coefficients. Dans ces deux cartes, les régions du sud-est de la France et de la Corse ont les coefficients les plus élevés ainsi qu'une partie de la Seine-et-Marne et de l'Alsace. Au contraire, les valeurs faibles des coefficients se situent plus dans le sud-ouest (jusqu'à la zone d'influence de Toulouse) et remonte jusqu'à Nantes le long de la façade atlantique.

L'effet simultanée de la défavorisation et de l'accessibilité aux médecins généralistes

L'ACP sur les résultats de la RGP montre que le premier axe de l'analyse porte la notion conjointe d'un impact de trois variables : *la défavorisation sociale, l'APL généraliste et l'accès santé*. Pour l'APL généraliste, seul les coefficients du troisième tercile par rapport au premier sont concernés. Les cartes présentées en annexes dans les figures 18 (APL généraliste), 22, 23 (Accès santé) et 24, 25 (FDep13) montre les territoires concernés par ces valeurs fortes des coefficients. Un large quart sud-est de la France regroupant les anciennes régions du Languedoc-Roussillon, de l'Auvergne, de Rhône-Alpes, de Provence-Alpes-Côte-D'azur et de Corse est concerné. Dans cette zone, l'impact de la défavorisation sociale, de l'accès à un médecin et une pharmacie en moins de 10 minutes et de l'accessibilité à un médecin généraliste y sont forts. Ce résultat offre un éclairage nouveau aux résultats inattendus de la

défavorisation sociale. L'impact de la défavorisation sociale semble étroitement lié à l'impact de l'accès aux médecins généralistes et non pas à celle des médecins gynécologues.

Partie V : Discussion

Les principaux résultats de l'étude peuvent être classés en deux catégories. La première catégorie porte sur les résultats de l'autocorrélation spatiale. La présence d'autocorrélation spatiale a été démontrée. Ce résultat relève d'une étape intermédiaire de la méthodologie de cette étude, mais constitue également un premier résultat important car il prouve la présence d'une disparité géographique dans le phénomène des taux de recours aux traitements de l'infertilité. Ce résultat correspond à nos hypothèses *a priori* basée sur les disparités de l'offre de soin en AMP en France (Bayer, et al. 2015) et la forte corrélation entre l'offre et le recours à ces soins d'AMP aux Etats-Unis (Hammoud, et al. 2009). Cependant, cette étude s'intéresse au recours des traitements de stimulation hormonale qui sont pris à toutes les étapes du parcours de soin de l'infertilité et non pas seulement au recours à l'AMP. De plus, il n'existe pas d'étude équivalente à celle de Hammoud et al en France. En Australie, l'article de Harris et al introduit l'aspect spatial sous la forme d'indicatrices de 7 grandes zones géographiques, mais ces variables semblent être surtout des variables de contrôle car les résultats associés à ces zones ne sont pas (ou très peu) commentés (Harris, et al. 2016). Le résultat de l'autocorrélation spatiale confirme la pertinence des analyses spatiales dans l'étude des taux de recours aux traitements de l'infertilité en France métropolitaine. Ce résultat devra être pris en compte dans les futures recherches sur le sujet.

La seconde catégorie concerne les résultats de la RGP. Le résultat le plus important tiré du RGP est celui portant sur l'APL pour les médecins gynécologues. La bonne accessibilité aux gynécologues favorise, partout en France métropolitaine, le recours aux traitements de l'infertilité. Ce résultat reflète probablement l'importance de la place de ces praticiens dans le parcours de soin des femmes, mais à notre connaissance, ce travail est le premier à objectiver cet effet. Etant donné l'impact des médecins gynécologues estimé dans cette étude, la fracture territoriale pour l'accès à cette spécialité (Figure 13) est inquiétante.

Les résultats de l'indice de défavorisation sociale FDep13 sont eux plus surprenants. En Australie, un indice similaire (*Socio Economic Index For Areas*, SEIFA) de mesure socioéconomique produit des résultats opposés (Harris, et al. 2016). Cet indice australien est également une agrégation au niveau d'une aire géographique de données individuelles donc il est comparable dans sa construction au FDep13 utilisé dans ce travail. En revanche, l'étude portant sur l'Australie ne mesure pas l'accès aux professionnels de santé et il est possible que cela soit un facteur de confusion au vu du lien entre les zones défavorisées et pauvres en médecins gynécologues. Il faut également noter que le système de remboursement français des traitements de l'infertilité est une spécificité qui pourrait impacter fortement les résultats. En Australie, le coût d'un cycle de transfert d'embryon supporté par le couple est compris entre 2 000 et 3 000 dollars américains (Harris, et al. 2016). En France, les traitements sont remboursés à 100 %.

5.1 Limites de l'étude

À propos du Modifiable Areal Unit Problem : MAUP

Le Modifiable Areal Unit Problem (MAUP) peut se traduire en français par Problème d'Agrégation Spatiale. Ce problème se pose dès lors que l'étude porte sur un découpage spatial, comme c'est le cas de cette étude avec le choix du territoire de vie. Le MAUP regroupe deux effets : l'effet de l'échelle et l'effet de zone (ESRI s.d.). L'effet échelle consiste à dire que la granularité de l'échelle utilisée détermine en partie les résultats de l'étude. Dans ce rapport, le choix de l'échelle intermédiaire a été développé dans la partie « 2.3 Choix de l'échelle d'agrégation spatiale ». Le territoire de vie, par sa

définition, est une échelle cohérente pour la question de l'étude mais il n'est pas exclu qu'une échelle différente pourrait amener à des résultats différents. L'effet de l'échelle dans les analyses multivariées est particulièrement examiné dans l'article de 1991 de Fotheringham (Fotheringham et Wong 1991). L'effet de zone est différent car il ne concerne pas le nombre d'unités créées par le découpage mais, à granularité fixée, l'effet de la forme des zones. Dans leur article, Fotheringham et Wong étudient les variations des estimations des coefficients de régressions linéaires et de régressions linéaires généralisées (logit) et concluent que le problème est « principalement imprédictible dans son intensité et son effet dans les analyses multivariées et dans ce sens un plus grand problème que dans les analyses univariées ou bivariées » (Fotheringham et Wong 1991). Ils soulignent également que l'effet de l'échelle est probablement plus intense car « l'estimation ne semble pas être aussi sensible à l'effet de zone qu'elle ne l'est à l'effet d'agrégation » (Fotheringham et Wong 1991). Certains auteurs, un peu plus optimistes, considèrent que le problème de l'agrégation spatiale peut être vu comme « une source d'information statistique nécessaire à la compréhension de la structure et des dynamiques des phénomènes complexes du monde réel » (Jelinski et Jianguo 1996). Dans le cas de cette étude, reproduire l'analyse avec une granularité différente permettrait de s'assurer de la robustesse des résultats.

La mesure de la distance aux centres AMP

Une des limites de l'étude est la prise en compte de l'activité des centres d'assistance médicale à la procréation. Dans l'étude, une notion de distance est introduite à l'aide de la contiguïté des territoires de vie. Chaque territoire peut posséder un centre AMP, être en première couronne d'un territoire de vie ayant un centre AMP ou en seconde couronne. Cependant, le rayonnement d'un centre AMP peut s'étendre très au-delà de deux territoires de vie adjacents. Notamment pour des métropoles comme Paris, Lyon et Marseille dont les arrondissements sont des territoires de vie dans cette étude. Une façon de palier à ce problème serait de considérer la distance entre la commune de résidence de la femme et le centre AMP le plus proche. Dans la pratique, ce type de distancier a été utilisé par l'équipe de recherche StimHo (Ben Messaoud, Bouyer et La Rochebrochard 2020) en collaboration avec un géographe de l'agence de biomédecine (ABM). Il s'appuie sur une moyenne du temps de trajet en voiture entre l'heure de pointe et l'heure creuse. Cependant, ces distanciers sont calculés pour les distances entre communes et il n'a donc pas pu être réutilisé ou recalculé dans cette étude qui s'appuie sur le niveau des territoires de vie. Le potentiel d'amélioration pour cette variable est donc important, notamment pour ce type d'étude géographique.

Quelles sont les autres pistes d'exploration possibles ?

Par ailleurs, cette étude est centrée sur l'environnement des patientes ayant recours aux traitements de l'infécondité. Analyser le profil des prescripteurs de ces traitements constitue une piste de recherche intéressante pour compléter cette recherche. Une première hypothèse est que le sexe et l'âge des prescripteurs présents dans un territoire de vie peut avoir une influence sur le taux de recours aux traitements de l'infécondité. Les données sur le nombre de médecins généralistes et gynécologues par tranche d'âge et par sexe ont fait l'objet d'une demande à la Drees le 26 mars 2020, qui n'a toutefois pas abouti avant la fin du stage du fait d'un manque de disponibilité des services de la Drees dans un contexte pandémique. Une autre caractéristique des prescripteurs à prendre en compte pourrait être leur secteur d'activité. Pour rappel, un médecin de secteur 1 ne pratique pas de dépassements d'honoraires, quand un médecin de secteur 2 peut en pratiquer. En complément, la proportion de médecins de secteur 2 OPTAM peut être un moyen d'affiner l'analyse. L'OPTAM est un contrat entre les médecins et l'Assurance maladie afin de maîtriser les dépassements d'honoraires. Enfin, la couverture maladie universelle complémentaire (CMU-C), disponible au niveau individuel dans le SNDS, pourrait être une variable intéressante à prendre en compte, dans la mesure où elle

constitue un proxy du niveau de ressources des individus. En effet, la CMU-C est une complémentaire santé pour les individus ayant de faibles ressources. Les conditions d'accès particulières au SNDS dans le contexte pandémique n'ont pas permis d'exploiter cette variable, mais l'intégrer à l'étude permettrait de compléter l'analyse de l'impact de la favorisation sociale en complément de la variable contextuelle.

Variables composites et choix des découpages en catégories

Les variables de défavorisation sociale, APL pour les généralistes et APL pour les gynécologues sont des variables composites car ce sont « des combinaisons mathématiques de plusieurs indicateurs qui représentent différentes dimensions d'un même concept » (Couralet, et al. 2010). Par exemple, l'indice de défavorisation sociale, FDep13, est une combinaison linéaire du taux de chômage, de la part d'ouvrier dans la population, de la part de bacheliers dans la population et du revenu médian. Bien que le développement des indicateurs composites se soit accéléré au cours des dernières décennies, ils sont controversés car ils « ne répondent pas à une méthodologie unique » et qu'il « n'existe pas de gold standard » (Couralet, et al. 2010). Certains résultats principalement obtenus à partir de mesures composites sont décriés, comme le classement des hôpitaux anglais à partir d'un indicateur composite de performance par Jacobs et Goddard qui mettent en question sa robustesse (Jacobs, Goddard et Smith 2005). Les indicateurs composites peuvent également poser un problème d'interprétation car ils n'ont pas nécessairement d'unité, à l'instar de la défavorisation sociale. Autant il est facile de se représenter le taux de chômage ou la part d'ouvrier dans une population, autant une combinaison de plusieurs variables peut perdre en lisibilité. Cependant, ils sont populaires pour leurs aspects synthétiques et pratiques lorsqu'il s'agit de comparer deux résultats. Pour le cas de la défavorisation sociale, une autre variable composite développée en France existe : *The European Deprivation Index* (EDI). Tester cette variable permettrait de tester la robustesse des résultats de la défavorisation sociale obtenus à l'aide du FDep13.

5.2 Forces de l'étude

Cette étude s'appuie sur les données du SNDS, qui rassemble de manière exhaustive les données de remboursement de soins, ce qui en fait une source très précieuse pour étudier la consommation de soins en population générale. En tant qu'organisme de recherche distingué par décret, l'Ined dispose d'un accès permanent à ces données : l'exploitation est d'autant plus facilitée par l'environnement sécurisé et l'expertise qui y sont déployées. La revue de littérature montre que le nombre d'étude portant sur tous les traitements de l'infertilité est faible en France et même en Europe. Cela s'explique probablement en partie par la difficulté d'avoir des données sur cette thématique, aussi la possibilité d'utiliser des données de la qualité de celles du SNDS est une force du projet de recherche StimHo dans lequel ce document s'inscrit.

A notre connaissance, aucune étude n'analyse l'accès aux traitements de l'infertilité avec un niveau de granularité spatial aussi fin que celui des territoires de vie pour la France métropolitaine. L'étude de Hammoud et al utilise le niveau des états aux Etats-Unis (Hammoud, et al. 2009). Harris et al utilise des indicatrices pour des régions australiennes (Harris, et al. 2016). Cette étude est la première étude géographique de l'accès aux traitements de l'infertilité sur un territoire national.

Enfin, les méthodes statistiques utilisées sont innovantes. En particulier, la régression géographiquement pondérée (RGP) est nouvelle et relativement simple : il s'agit d'exécuter un grand nombre de régressions pour chaque unité de l'espace. La puissance de calcul importante à disposition aujourd'hui permet son utilisation, relativement gourmande en ressources. Les modèles RGP sont de plus en plus recommandés pour l'analyse spatiale.

Conclusion et enseignements

Cette étude montre l'importance de l'aspect spatial dans l'analyse du recours aux traitements de l'infertilité pour les femmes en France métropolitaine. Dans cette étude, nous nous sommes concentrés sur l'étude de l'hétérogénéité spatiale en réalisant un modèle RGP. Cependant, la dépendance spatiale peut prendre la forme d'une hétérogénéité spatiale ou d'une autocorrélation spatiale. Ces deux formes ne sont pas incompatibles l'une de l'autre et il est possible que l'une influe sur l'autre. Notamment, il n'existe pas de tests statistiques permettant de décider s'il est préférable de faire l'hypothèse d'une hétérogénéité ou d'une autocorrélation spatiale. Dans la perspectives de futures recherches sur le sujet, je conseillerais de chercher à étudier également la notion d'autocorrélation spatiale. Dans le cas de régression linéaire simple, cela passerait par la mise en place de modèle spatialement autorégressif (SAR), de modèle de Durbin-Watson (SDM), modèle d'erreur spatiale (SEM) ou de modèle à retards spatiaux (SLX). Pour choisir entre les différents modèles, des stratégies de tests peuvent être mis en place dont la plus populaire est sans doute la méthode d'Elhorst (Elhorst 2010). Cependant, cette méthode de test n'est valable que pour des régressions de moindre carré ordinaire (MCO) et non pas pour des modèles de Poisson. Dans le cas de modèles de Poisson, ou plus généralement de modèle GLM, je conseille de faire un modèle spatial linéaire généralisé à effets mixtes (SGLMM, Spatial generalised linear mixed models). « Dans le cas de données spatiales, les GLMMs peuvent être utilisés pour estimer les effets fixes en contrôlant l'autocorrélation spatiale entre les régions » (Dormann, et al. 2007).

Un autre résultat de cette étude est le rôle important des médecins gynécologues dans le recours aux traitements de l'infertilité pour les femmes. Cet impact n'est pas surprenant au vu de l'importance de ces médecins dans le parcours de soins des patientes. A la lumière de ce résultat, la fracture territoriale dans l'accès aux médecins gynécologues en France est inquiétante. Une politique publique ayant pour objectif de permettre un égal accès aux traitements de l'infertilité devrait vraisemblablement se concentrer sur les disparités existantes en matière d'accès aux médecins gynécologues.

L'équipe de recherche StimHo n'avait jusqu'alors pas analysé les données sur le nombre d'acheteuses de traitements de l'infertilité sous l'angle spatial. Ce document est donc un premier travail et dans ce sens il peut répondre à la question « Est-ce une voie sans issues ou bien un terrain fertile de recherche ? ». A mon sens, les résultats de ce travail montrent que ça n'est pas une voie sans issues. La présence d'autocorrélation spatiale indique la nécessité de prendre en compte cet aspect. Le résultat attendu de l'impact des médecins gynécologues peut ouvrir à de prochaines investigations (profils démographiques des prescripteurs ?) et le résultat inattendu de la défavorisation sociale montre que ce point nécessite de futures recherches.

Bibliographie

- Adashi, Eli Y, et Laura A Dean. 2016. «Acces to and use of infertility services in the united states : framing the challenges.» Édité par Elsevier Inc. *American Society for Reproductive Medecine* 105 (5): 1114-1118.
- Ameli. 2010. *La prise en charge de l'infertilité*. 10 Juillet. Accès le Juillet 28, 2020. <https://www.ameli.fr/assure/sante/themes/sterilite-pma-infertilite/prise-charge-infertilite>.
- Barlet, Muriel, Magali Coldefy, Clémentine Collin, et Véronique Lucas-Gabrielli. 2012. *L'Accessibilité potentielle localisée (APL) : une nouvelle mesure de l'accessibilité aux soins appliquée aux médecins généralistes libéraux en France*. Drees.
- Bayer, Florian, Sabrina Di Costanzo, Françoise Merlet, Fabienne Pessionne, et Dominique Royere. 2015. *Atlas de l'offre de soins en assistance médicale à la procréation en France*. ABM.
- Bellefon, Marie-Pierre, et Jean-Michel Floch. 2018. «Chapitre 9 - Régression géographiquement pondérée.» Dans *Manuel d'analyse spatiale. Théorie et mise en oeuvre pratique avec R*, de Insee - Eurostat. Sous la direction de Vincent Loonis. Coordonné par Marie-Pierre de Bellefon, 240-260. Insee méthodes.
- Bellefon, Marie-Pierre, Vincent Loonis, et Ronan Le Gleut. 2018. «Codifier la structure de voisinage.» Chap. 2 dans *Manuel d'analyse spatiale - Théorie et mise en oeuvre pratique avec R*, de Vincent Loonis et Marie-Pierre Bellefon, 34-48. Insee.
- Ben Messaoud, Khaoula. 2018. «Epidemiologie des traitements de l'infécondité par stimulation hormonale explorés avec les données du Sniiram dans la cohorte Constance et l'EGB.» Université Paris-Saclay, Ined et StimHo, 23 Juin.
- Ben Messaoud, Khaoula, Jean Bouyer, et Elise (de) La Rochebrohard. 2020. «Infertility treatment in France, 2008-2017 : a challenge of growing treatment needs at older ages.» *American Journey of Public Health*.
- Boivin, Jacky, Laura Bunting, John A. Collins, et Karl G. Nygren. 2007. «International estimates of infertility prevalence and treatment-seeking : potential need and demand for infertility medical care.» *Human Reproduction* 22 (6): 1506-1512.
- Breuil-Genier, Pascale, et Céline Goffette. 2006. *La durée des séances des médecins généralistes*. Drees.
- Brunsdon, Chris, Steward Fotheringham, et Martin Charlton. 1996. «Geographically weighted regression : a method for exploring spatial nonstationarity.» *Geographical analysis* 281-298.
- Carstairs, V, et R Morris. 1989. «Deprivation : explaining differences in mortality between Scotland and England and Wales.» *Bmj* 886-889.
- Chandra, A, CE Copen, et EH Stephen. 2014. «Infertility service use in the United States : data from the national survey of family growth, 1982-2010.» *Natl Health Stat Rep* (73): 1-21.
- Couralet, Mélanie, Sophie Guérin, Marc Le Vaillant, Philippe Loirat, et Etienne Minvielle. 2010. *Analyse critique du développement d'indicateurs composites : le cas de l'infarctus du myocarde à la sortie de l'établissement de santé*. Rapport, Compaqh - Inserm.

- Daar, Abdallah S, et Zara Merali. 2001. «Infertility and social suffering : the case of ART in developing countries.» Dans *Current practices and controversies in assisted reproduction.*, de E Vayena, P.J Rowe et Griffin P.D, 15-21. Geneva: World Health Organisation.
- Davis, Owen K., et Rebecca Z. Sokol. 2016. «Acces to fertility care.» *American Society for Reproductive Medecine* 105 (5): 1111-1112.
- de La Rochebrochard, Elise. 2004. «Médicalisation de l'infertilité : quelle est la situation mondiale du Nord au Sud ?» *Santé de la reproduction au Nord et au Sud : de la connaissance à l'action* 277-292.
- DHHS. 2019. *About the National Survey of Family Growth*. 4 Janvier. Accès le Juillet 21, 2020. https://www.csc.gov/nchs/nsfg/about_nsfg.htm.
- Domar, Alice D., Kristin Smith, Lisa Conboy, Marie Iannone, et Michael Alper. 2010. «A prospective investigation into the reasons why insured United States patients drop out of in vitro fertilization treatment.» *Fertility and Sterility* 94 (4): 1457-1459.
- Dormann, Carsten, Jana McPherson, Miguel Araujo, Roger Bivand, Janine Bolliger, Gudrun Carl, Richard Davies, et al. 2007. «Methods to account for spatial autocorrelation in the analysis of species distributional data : a review.» *Ecography* 609-628.
- Elhorst, J.P. 2010. «Applied Spatial Econometrics : Raising the Bar .» *Spatial Economic Analysis* 9-28.
- ESRI. s.d. *esri - support technique*. Accès le Juillet 23, 2020. <https://support.esri.com/fr/other-resources/gis-dictionary/search/>.
- Fergani, Nasreddine, et Ameer Soukhal. 2015. «Algorithme incrémental pour les graphes des voisins relatifs.» Édité par hal-01135748. *ROADEF*.
- Fotheringham, AS, et DWS Wong. 1991. «The modifiable areal unit problem in multivariate statistical analysis.» *Environment and Planning* 23: 1025-1044.
- Hammoud, Ahmad, Mark Gibson, Joseph Stanford, George White, Douglas Carrell, et Matthew Peterson. 2009. «In vitro fertilization availability and utilization in the United States : a study of demographic, social and economic factors.» *Fertility and Sterility* 91 (5): 1630-1634.
- Harris, Katie, Hugh Burley, Robert McLachlan, Mark Bowman, Alan Macalodowie, Kate Taylor, Michael Chapman, et Georgina Mary Chambers. 2016. «Socioeconomic disparities in acces to assisted reproductive technologies in Australia.» *Reproductive Healthcare Ltd*. 33: 575-584.
- Harvard, Sabrina. 2008. «A small-area index of socioeconomic deprivation to capture health inequalities in France.» *Elsevier Ltd*.
- Hypothèses, StimHo - Carnet de recherche proposé par. s.d. *Le projet StimHo*. Accès le Juillet 20, 2020. <https://stimho.hypotheses.org/presentation/projet-stimho>.
- Jacobs, Rowena, Maria Goddard, et Peter Smith. 2005. «How robust are hospital ranks based on composite performance measures.» *MedCare* 1177-1184.
- Jain, Tarun. 2006. «Socioeconomic and racial disparities among infertility patients seeking care.» *Fertility and Sterility* 85 (4): 876-881.
- Jelinski, Dennis, et Wu Jianguo. 1996. «The modifiable areal unit problem and implications for landscape ecology.» *Landscape Ecology* 11 (3): 129-140.

- Lasbeur, Linda. 2006. «Analyser les inégalités socioéconomiques de santé à partir des données du recensement.» *Institut national d'études démographiques - Population* 61 (4): 567-584. <https://www.cairn.info/revue-population-2006-4-page567.htm>.
- Le Gallo, Julie. 2004. «Hétérogénéité spatiale : principes et méthodes.» *Economie et prévision* 1 (162): 151-172. doi:10.3406/ecop.2004.6939.
- Lucas-Gabrielli, Véronique, Clément Nestrigue, et Magali Coldefy. 2016. *Analyse de sensibilité de l'Accessibilité Potentielle Localisée*. Document de travail, Irdes.
- Morris, Melanie, Laura Oakley, Noreen Maconochie, et Pat Doyle. 2011. «An investigation of social inequalities in help seeking and use of health services for fertility problems in a population based sample of UK women.» *Human Fertility* 14 (1): 16-22.
- Nakaya, T, A.S Fotheringham, C Brunson, et M Charlton. 2005. «Geographically weighted Poisson regression for disease association mapping.» *Statistics in medicine* 2695-2717.
- O'Brien, Robert. 2007. «A Caution Regarding Rules of Thumb for Variance Inflation Factors.» *Quality and Quantity* 41: 673-690.
- Quinn, Molly, et Victor Fujimoto. 2016. «Racial and ethnic disparities in assisted reproductive technology acces and outcomes.» *American Society for Reproductive Medecine* 1119-1123.
- Rey, Grégoire. 2007. «Surmortalité lié aux vagues de chaleur : facteurs de vulnérabilité sociodémographiques et causes médicales de décès.» Université Paris XI.
- Rey, Grégoire, Eric Jouglu, Anne Fouillet, et Denis Hémon. 2009. «Ecological association between a deprivation index and mortality in france over the period 1997-2001 : variations with spatial scale, degree of urbanicity, age, gender and cause of death.» *BMC Public Health* 33 (9): 1-12.
- Reynard, Robert, et Pascal Vialette. 2014. *Une approche de la qualité de vie dans les territoires*. 08 Octobre. Accès le Juillet 10, 2020. <https://www.insee.fr/fr/statistiques/1281328#documentation>.
- Salima, Bouayad Agha, et (De) Marie-Pierre Bellefon. 2018. «Indices d'autocorrélation spatiale.» Chap. 3 dans *Manuel d'analyse spatiale*, de Vincent Loonis et (De) Marie-pierre Bellefon, 54-70. Insee.
- Smith, James, Michael Eisenberg, David Glidden, Susan Millstein, Marcelle Cedars, Thomas Walsh, Jonathan Showstack, Lauri Pasch, Nancy Adler, et Patricia Katz. 2011. «Socioeconomic disparities in use and succes of fertility treatment : analysis of data from a prospective cohort in the united states.» *American Society for Reproductive Medecine* (Elsevier Inc) 96 (1): 95-101.
- Townsend, P. 1987. «Deprivation.» *Journal of Social Policy* 16 (2): 125-146.
- Vergier, Noémie, et Ingrid Lefebvre-Hoang. 2017. *Déserts médicaux : Comment les définir ? Comment les mesurer ?*. Drees.
- Vergier, Noémie, Hélène Chaput, et Ingrid Lefebvre-Hoang. 2017. *Désert médicaux : comment les définir ? Comment les mesurer*. Drees.
- White, Lynn, Julia McQuillan, et Arthur Greil. 2006. «Explaining disparities in treatment seeking : the case of infertility.» *Fertility and Sterility* 85 (4): 853-857.

WHO. 2010. *Mother or nothing : the agony of infertility*. Vol. 88, chez *Bulletin of the World Health Organization*, de WHO, 881-882.

World Health Organization. 2020. *Infertility is a global public issue*. Mai.
www.who.int/reproductivehealth/topics/infertility/perspective/en/.

Annexe A : Glossaire des variables explicatives

L'indice de défavorisation sociale

La favorisation sociale est un indice qui est ici utilisé pour estimer l'impact de la barrière économique. Dans sa construction, cet indicateur ressemble à celui utilisé par Harris dans son article de 2016 pour mesurer le statut socioéconomique des femmes « Index of Relative Socioeconomic Advantage and Disadvantage (ISRAD) » (Harris, et al. 2016).

Nommé FDep13, l'indicateur de désavantage (ou défavorisation) social FDep a été développé par Grégoire Rey et ses collègues en 2009 dans le but de « refléter au mieux le concept de défavorisation » et de présenter « les contrastes socioéconomiques les plus importants » (Rey, Jouglu, et al. 2009). Cet indicateur s'appuie sur les données de l'Insee du recensement de la population et des déclarations d'impôt. Il est construit au niveau communal sur l'ensemble du territoire. Une première version, le FDep09, est construit avec les données du recensement de 2009. Par la suite, de nouvelles versions ont vues le jour avec des données plus récentes. Dans cette étude, le FDep13 construit avec les données de 2013 est utilisé.

L'indicateur est la première composante principale d'une analyse en composante principale (ACP) portant sur les quatre variables suivantes définies au niveau communal :

- Revenus médians des ménages
- Part de la population (15 ans ou plus) avec le baccalauréat
- Part d'ouvriers dans la population
- Taux de chômage dans la population

Le choix de ces variables et la méthode pour construire l'indicateur s'appuie sur une littérature préexistante qui est plus détaillée dans la partie annexe de ce rapport.

L'indice d'accessibilité potentielle localisée (APL)

Les indicateurs d'accessibilité potentielle localisée (APL) servent à mesurer la tension entre l'offre et la demande de soins dans une région donnée. Il s'agit avec ces indicateurs de mettre en évidence une éventuelle barrière géographique, dans le sens où ils peuvent servir à comparer les territoires en matière d'accessibilité de soins.

L'APL est construit par la Drees et l'Irdes une première fois en 2012 dans sa version 0. Une version 1 et une version 2 ont également été produites respectivement en 2017 et 2020. Cet indicateur est défini au niveau de la commune et il s'exprime en nombre de consultations disponibles par habitant et par an (C/hab/an) depuis la version 1. L'APL aide notamment les Agences Régionales de Santé (ARS) à identifier les zones de déserts médicaux. Dans le cas de cette étude, il concerne l'offre de soin des médecins généralistes et gynécologues.

- Pour les médecins généralistes, il sera dénommé **apl_gen**
- Pour les médecins gynécologues, il sera dénommé **apl_gyn**

Pour les gynécologues, il a été calculé lors de la première version en 2012 mais n'a pas reçu de mise à jour depuis. Le paramétrage exact pour l'APL associé aux gynécologues n'est plus disponible sur le site de la Drees. Une annexe est dédiée à la méthodologie précise de ces indicateurs APL.

Le taux d'urbanisation

Le taux d'urbanisation du territoire de vie est calculé à partir des données produites en France par le service statistique ministériel du ministère en charge de l'écologie en collaboration avec, depuis 2018, l'Institut national de l'information géographique et forestière (IGN). Elles sont disponibles sur le site <https://www.statistiques.developpement-durable.gouv.fr/corine-land-cover-0>. La méthode consiste à découper des images satellitaires en parcelles d'environ 25 ha, puis de classer ces parcelles dans la nomenclature. Cette nomenclature comporte trois niveaux de finesse, le niveau le plus fin ayant 44 postes. Il existe cinq grands types de territoires :

- Territoires artificialisés
- Territoires agricoles
- Forêts et milieux semi-naturels
- Zones humides
- Surfaces en eau

Pour cette étude, le taux d'urbanisation correspond à la proportion de territoires artificialisés sur la surface totale du territoire de vie pour l'année 2018.

L'activité des centres AMP

L'activité des centres d'AMP est mesurée par le nombre de prélèvements d'ovocytes qui y sont effectués au cours d'une année, ici 2015. Cette activité est normalisée par le nombre de femmes âgées de 18 à 49 ans du territoire de vie. L'unité de mesure est un nombre de ponctions d'ovocytes pour 10 000 femmes. Cet indicateur a été calculé à l'aide de données extraites à partir du rapport annuel de l'ABM pour l'année 2015 : <https://www.agence-biomedecine.fr/annexes/bilan2016/donnees/procreation/01-amp/synthese.htm>. "La participation de ces centres est une obligation réglementaire depuis 2010", ce qui assure une qualité des données "malgré les nombreuses difficultés de terrain".

Il faut noter que l'activité des centres AMP sur le territoire de vie ne reflète pas le nombre de ponctions d'ovocytes réalisé sur des femmes de ce territoire de vie. En effet, les centres peuvent avoir un rayonnement au-delà du territoire de vie auquel ils appartiennent, parfois même au-delà de la région.

Distance aux centres AMP

Pour mesurer plus précisément l'impact de la barrière aux infrastructures, il est nécessaire de disposer d'une variable qui caractérise la distance aux centres AMP. La construction de cette variable s'appuie sur une matrice de contiguïté. La variable se décompose en quatre catégories :

- Le TV dispose d'un centre d'AMP en son sein.
- Le centre d'AMP le plus proche du TV se situe en première couronne du TV.
- Le centre d'AMP le plus proche du TV se situe en deuxième couronne du TV.
- Le centre d'AMP le plus proche du TV se situe au-delà de la deuxième couronne du TV.

Accès santé

Cette variable représente la part (en %) de la population du territoire de vie ayant un accès en moins de dix minutes à un médecin généraliste et à une pharmacie. Les informations proviennent de l'Insee à partir de la base permanente des équipements et du recensement de la population. La distance est produite sur la base de METRIC, fournie par l'Insee.

Cette variable a une distribution particulière car 62,5 % des territoires de vie ont 100 % ayant un accès en moins de dix minutes à une pharmacie et à un généraliste. La variable sera par la suite découpée en trois catégories.

- Les territoires de vie à 100 % forment une catégorie
- Les territoires de vie entre 90 et 99 %
- Les territoires de vie en dessous de 90 %

Annexe B : Les cartes des variables FDep, APL, Urbanisation, Activités des centres et accès santé

Figure 10 : Distribution du FDep13 en France

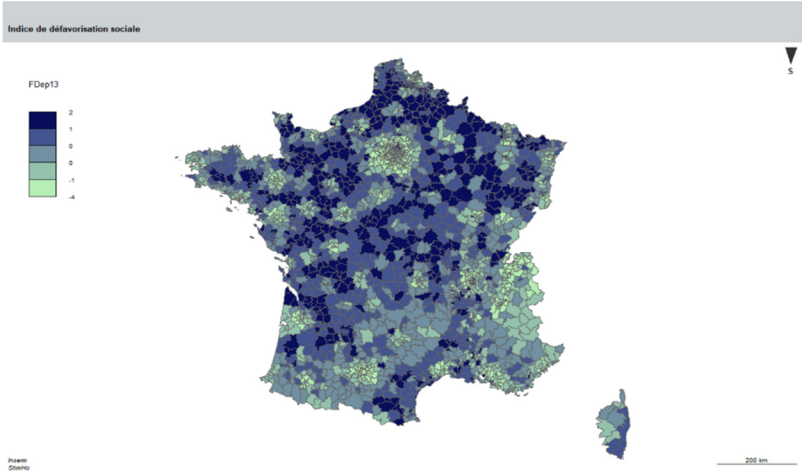
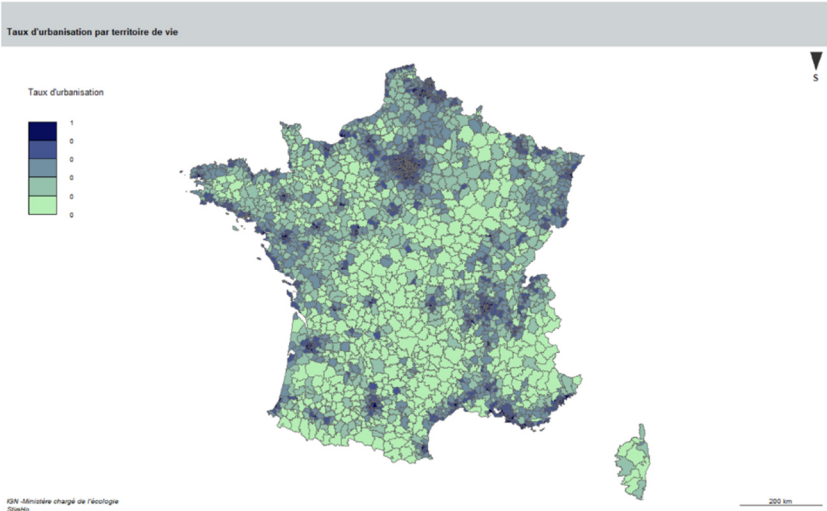


Figure 11 : Taux d'urbanisation des territoires de vie en France métropolitaine



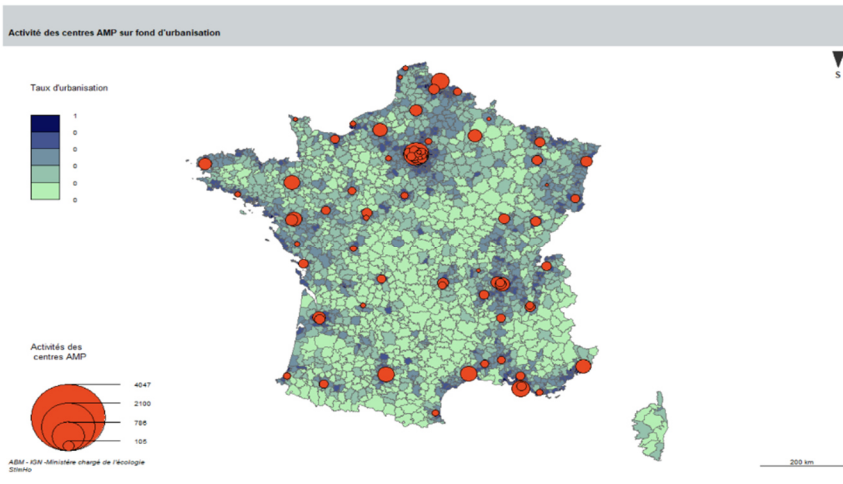


Figure 12 : Localisation et activités des centres AMP en France métropolitaine (fond de carte taux urbanisation)

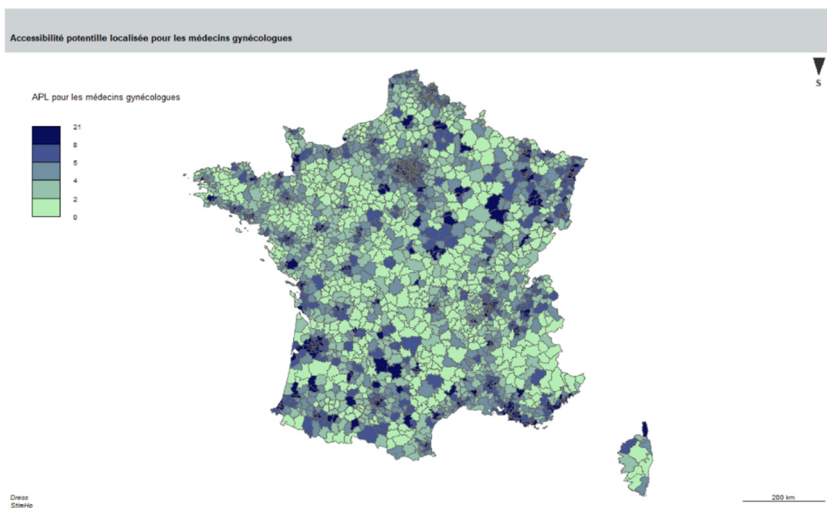
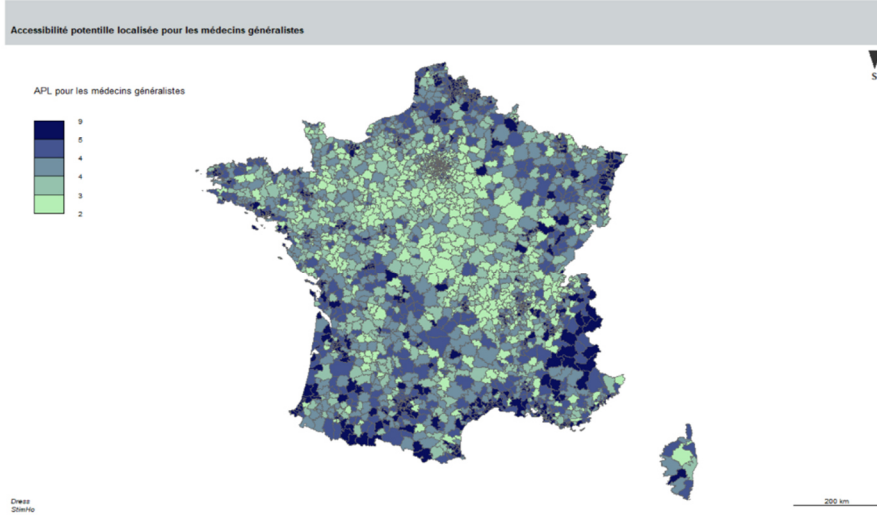


Figure 13 : Répartition de l'APL gynécologue sur le territoire de France

Figure 14 : Répartition de l'APL généraliste en France



Annexe C : Cartes RGP – APL pour les médecins gynécologues

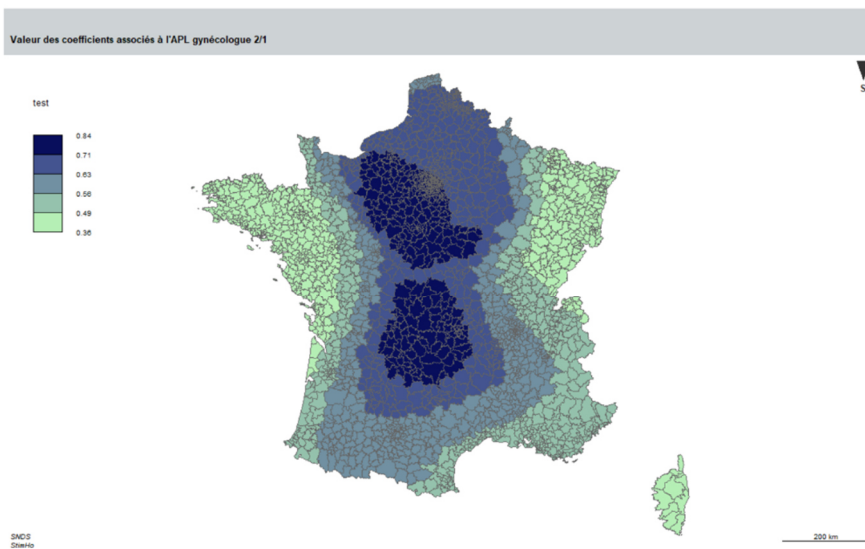


Figure 15 : Estimations de la RGP des coefficients associées à l'APL gynécologue 2/1

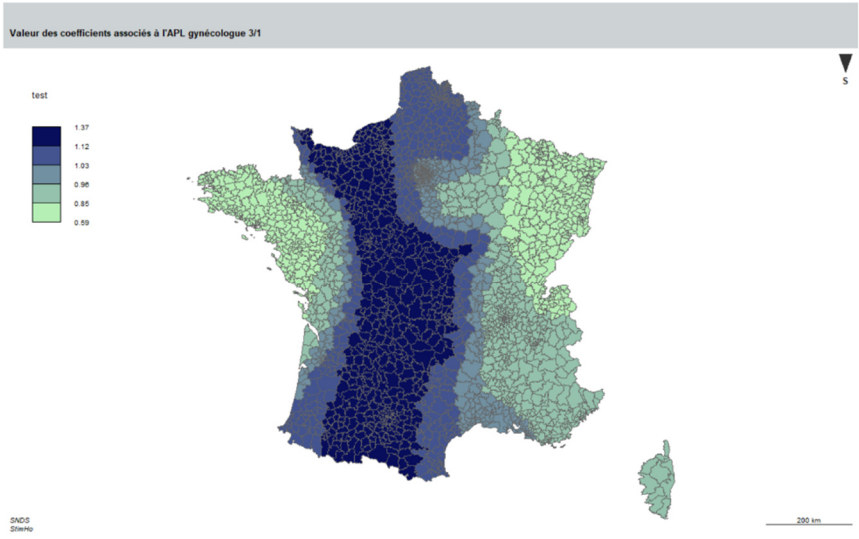


Figure 16 : Estimations de la RGP des coefficients associées à l'APL gynécologue 3/1

Annexe D : Cartes RGP – APL pour les médecins généralistes

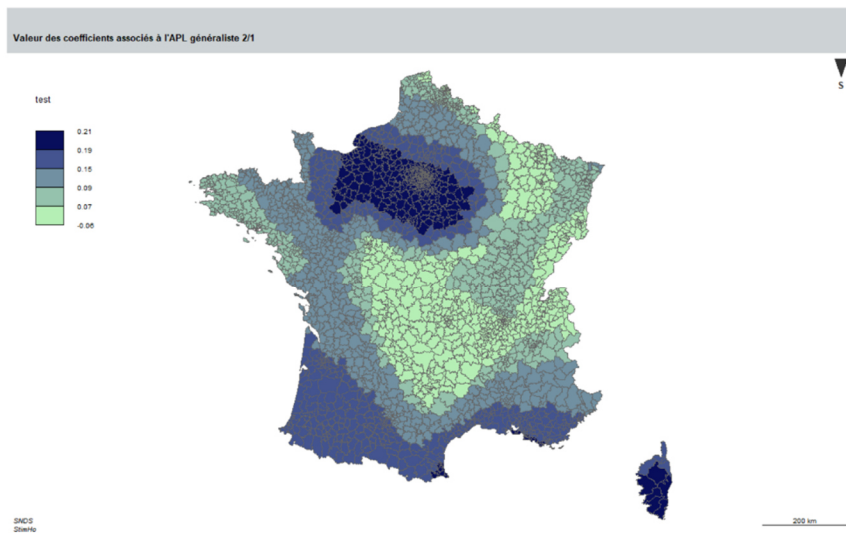


Figure 17 : Carte des coefficients de la RGP pour l'APL généraliste (2/1)

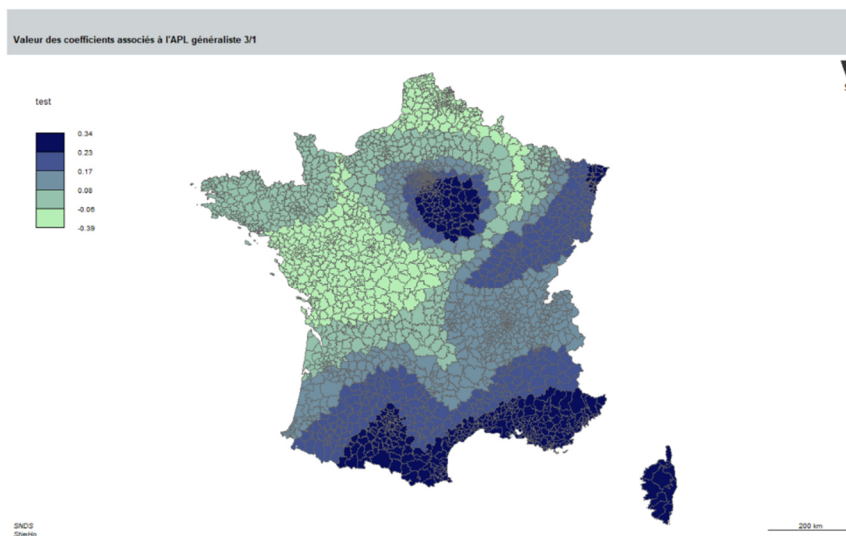


Figure 18 : Carte des coefficients de la RGP pour l'APL généraliste (3/1)

Annexe E : Carte RGP – Distance aux centres d'AMP

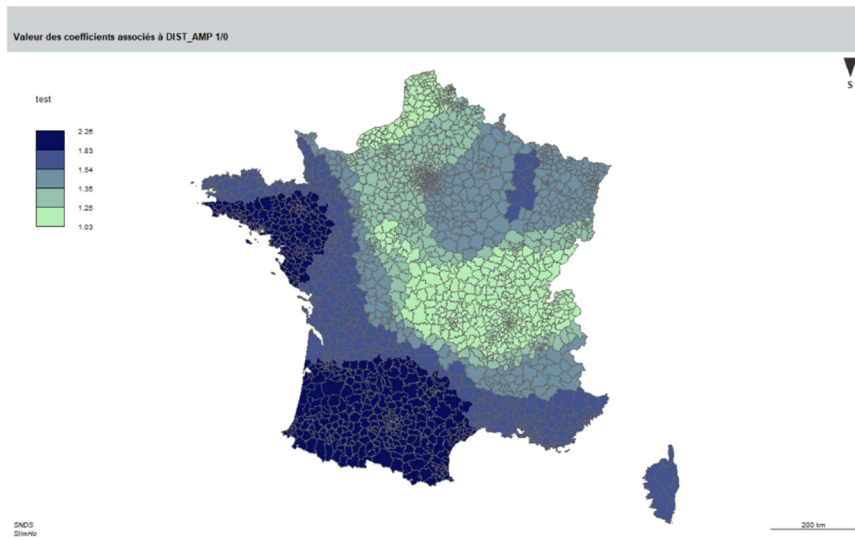


Figure 19 : Estimations de la RGP des coefficients associées à la distance aux centres AMP (1/0)

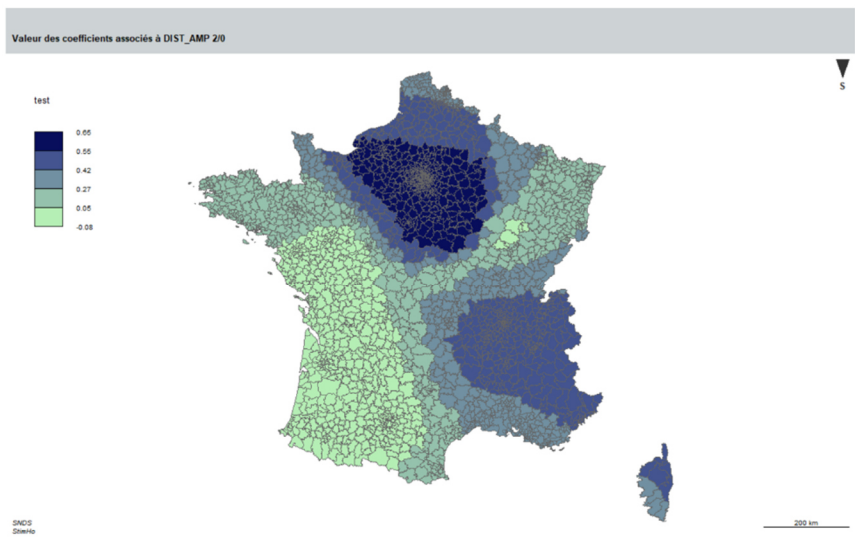


Figure 20 : Estimations de la RGP des coefficients associées à la distance aux centres AMP (2/0)

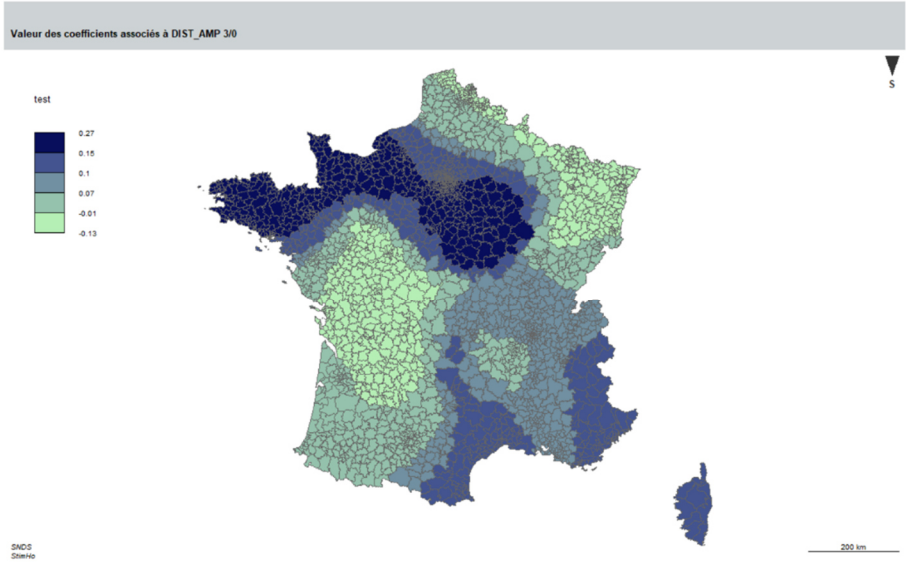


Figure 21 : Estimations de la RGP des coefficients associées à la distance aux centres AMP (3/0)

Annexe F: Cartes RGP – L'accès aux services de santé de bases

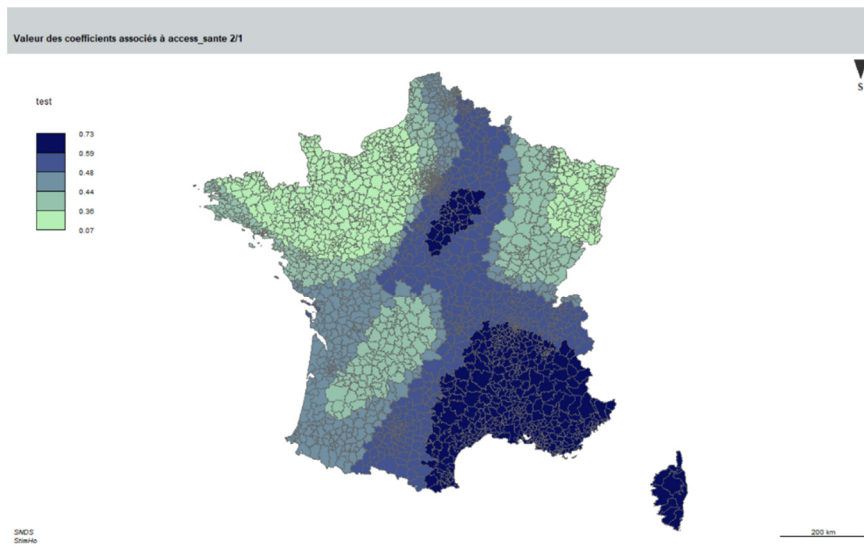


Figure 22 : Estimations de la RGP des coefficients associées à l'accès aux services de santé (2/1)

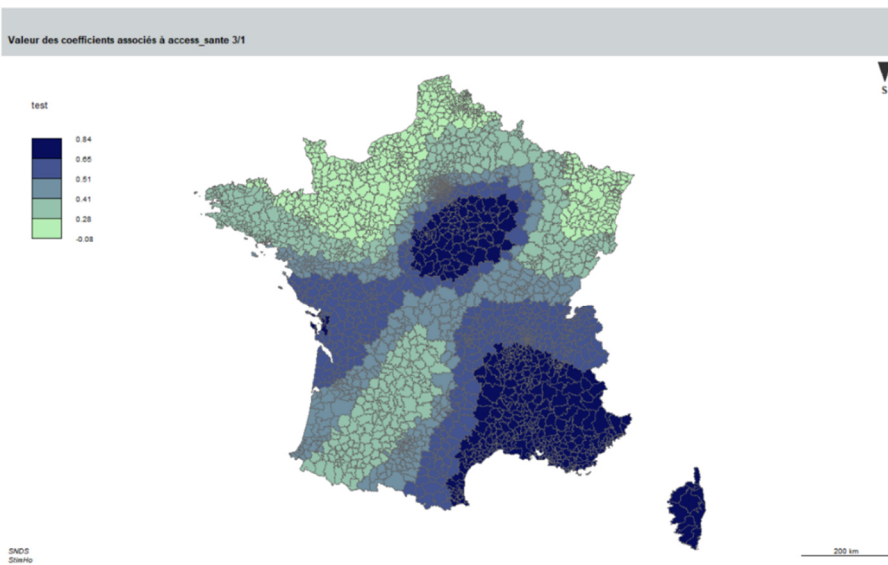


Figure 23 : Estimations de la RGP des coefficients associées à l'accès aux services de santé (3/1)

Annexe G : Carte RGP – L'indice de défavorisation sociale

Figure 24 : Carte des coefficients de la RGP pour le FDep13 (2/1)

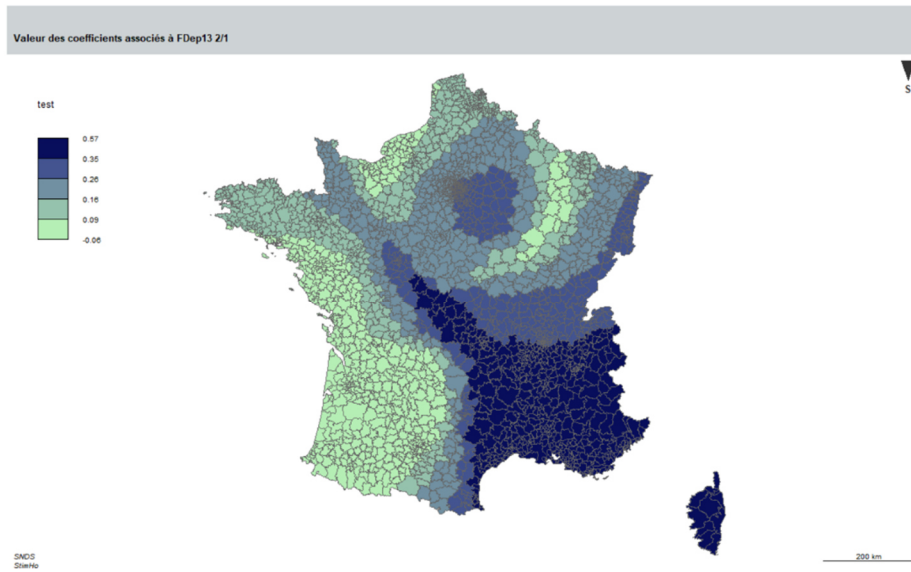


Figure 25 : Carte des coefficients de la RGP pour le FDep13 (3/1)

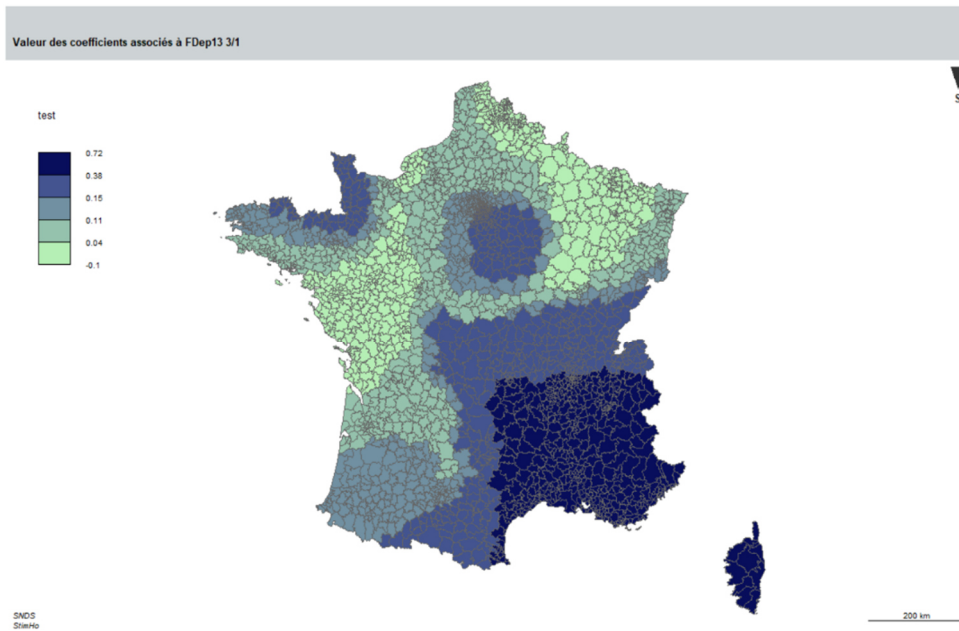


Tableau 8 : Proportion de la population de femmes âgées de 18 à 49 ans vivant dans les différents quintiles des variables de l'étude

Quintile	1	2	3	4	5
Taux de recours	10.8	17.9	19.89	20.88	30.52
FDep13	24.63	23.16	19.52	16	16.69
Apl_gyn	8.47	13.33	17.08	20.12	41.01
Apl_gen	16.25	15.88	16.82	21.71	29.35
Urbanisation	8.17	13.75	18.25	18.98	40.85

Note de lecture : 17,9 % des femmes âgées de 18 à 49 ans habitent dans un territoire de vie appartenant au deuxième quintile de la distribution des taux de recours