

# **Etude (étiologique) des facteurs associés à un stade diagnostic évolué du cancer du sein chez la femme en région Corse entre 2017 et 2019**

Yannick VIDAL<sup>1</sup>

<sup>1</sup> *Registre général des cancers de Corse (RG2C), Observatoire Régional de la Santé de Corse (ORS-C), Ajaccio, France*

## **Résumé**

**Introduction :** L'hypothèse formulée dans cette étude est que la probabilité de diagnostic d'un cancer du sein à un stade « évolué » est influencée par des facteurs liés à la maladie, au patient et à son environnement. L'objectif principal de l'étude est l'identification des facteurs associés à un diagnostic à un stade « évolué » du cancer.

**Méthodologie :** Une étude étiologique transversale rétrospective à partir des données du RG2C enrichie avec des données disponibles en opendata a été mise en place. Un modèle logistique a été défini afin de quantifier les effets des différentes covariables sur la probabilité de diagnostic du cancer à un stade « évolué ».

**Résultats :** Seules les circonstances de la découverte ont un impact statistiquement significatif sur la probabilité de diagnostic d'un cancer à un stade évolué. L'augmentation du risque est de 2,06 IC95[1,27 ; 3,34] lors d'un diagnostic posé par le médecin et de 4,18 IC95[2,43 ; 7,18] lors de signes d'appel détectés par le patient vs le dépistage organisé.

**Discussion :** Ces résultats démontrent une nouvelle fois l'importance du dépistage organisé dans la détection du cancer du sein dès le stade « limité » même si les résultats nécessitent d'être confrontés à des études similaires dans des zones géographiques comparables.

## **Mots clés**

Tumeur du sein ; Corse ; RG2C ; Stade au diagnostic



## Riassuntu

**Intruduzione :** L'ipotesi formulata in 'ssu studiu hè chì a prubabilità di un diagnosticu di cancaru di u pettu in fasa « avanzata » sia influenzata da fattori ligati à a malatia, à u paziente è à u so ambiente di vita. U fine principale di u studiu hè d'identificà i fattori assuiciati à un diagnosticu in fasa « avanzata » di u cancaru.

**Metudulugia :** Hè stata fatta un'inchiesta etiologica trasversale retrospettiva à partesi da i dati di u RG2C, arricchiti cun dati disponibili in opendata. Un mudellu lugisticu hè statu definitu per quantificà l'effetti di e sfarente covariabile annantu à a prubabilità di diagnosticu di u cancaru in fasa « avanzata ».

**Risultati :** Solu e circustanze di a scuperta anu un impattu significativu da u puntu di vista statisticu annantu à a prubabilità di diagnosticu di un cancaru in fasa avanzata. U risicu cresce di 2,06 [1,27 ; 3,34] quandu u diagnosticu hè fattu da un medicu, è di 4,18 [2,43 ; 7,18] quandu si tratta di segni d'alerta rilivati da u paziente, per rapportu à a vistighera organizata.

**Discussion :** Sti risultati mettenu torna in risaltu l'impurtanza di a vistighera organizata in a ditezzione di u cancaru in un statu di sviluppu « limitatu », ancu s'ella ferma necessaria di cunfruntà i risultati cù inchieste simile fatte in zone geografiche paragunevule.

### *Parolle chjave*

Cancaru di u pettu ; Corsica ; RG2C ; statu di sviluppu à u diagnosticu

## Abstract

**Introduction:** The hypothesis of this study is that the probability of being diagnosed with breast cancer at an "advanced" stage is influenced by factors related to the disease, the patient, and their environment. The primary objective of the study is to identify factors associated with a diagnosis at an "advanced" stage of cancer.

**Methodology:** A retrospective cross-sectional etiological study was conducted using data from the RG2C database, enriched with open data sources. A logistic regression model was used to quantify the effects of various covariates on the probability of being diagnosed with breast cancer at an advanced stage.

**Results:** Only the circumstances of discovery had a statistically significant impact on the likelihood of an advanced-stage diagnosis. The risk increased by 2.06 [1.27; 3.34] when the diagnosis was made by a physician, and by 4.18 [2.43; 7.18] when symptoms were detected by the patient, compared to organized screening.

**Discussion:** These findings once again highlight the crucial role of organized screening in detecting breast cancer at a "limited" stage. However, the results should be compared with similar studies conducted in comparable geographic areas.

### *Keywords*

Breast tumor ; Corsica ; RG2C ; Stage at diagnosis



## Remerciements

Je souhaiterais en premier lieu remercier la personne qui a permis ma reprise d'études en inscrivant ma volonté personnelle au sein d'un projet global d'évolution de mon poste au sein de l'observatoire régional de la santé. Il s'agit de l'ancien directeur de la structure, M. Christophe ARENA. Merci d'avoir cru en mon potentiel et d'avoir mis en œuvre les moyens pour initier ma démarche.

Je tiens également à remercier l'actuelle directrice de l'ORS, Mme Estelle CERVETTI-CASTA, d'avoir prolongé la démarche engagée malgré les difficultés rencontrées par l'association durant ces deux dernières années. Merci d'avoir poursuivi l'effort engagé par M. ARENA en permettant la continuité de mon parcours universitaire.

Mes remerciements s'adressent également au Dr. Paul-André COLOMBANI, président de l'ORS et député de la deuxième circonscription de Corse-du-Sud, pour avoir donné un avis favorable à ma démarche et encouragé le conseil d'administration à soutenir mon projet de reprise d'études.

Également, je remercie les personnes m'ayant accompagné tout au long de ce stage en me prodiguant leurs précieux conseils médicaux, statistiques et méthodologiques.

Merci au Dr. Jean-Pierre AMOROS d'avoir accepté d'être mon tuteur de stage et de s'être impliqué dans toutes les étapes de mon parcours, depuis la préparation du sujet de stage à la relecture du rapport final.

Merci à M. Jean ARRIGHI pour son assistance durant ce stage et de m'avoir mis à disposition tout son savoir de biomathématicien.

Merci à M. Laurent LEHOT d'avoir accepté d'être mon tuteur pédagogique. Votre compétence et vos connaissances en statistiques ont été indéniablement un atout dans l'exploitation et l'interprétation de mes modèles et de mes données.

Je n'oublie pas non plus les personnes qui ont été disponibles quand cela a été nécessaire.

Merci à Mme Sandrine NOBLET pour la relecture attentive de ce rapport et pour les nombreux articles scientifiques communiqués en lien avec le cancer du sein et les stades aux diagnostics.

Merci à Mme Françoise GRAZIANI pour la traduction du résumé en langue corse.

Pour finir je souhaiterais remercier la personne sans qui ce travail n'aurait pas pu être mené à bien : Mme Laura ALESSANDRI. Merci pour ton soutien indéfectible et pour avoir pris soin de notre fils durant de longues heures les soirs et week-ends afin que je puisse rendre ce travail dans les délais.

## Table des matières

Résumé .....	1
Remerciements .....	5
Table des matières .....	6
Glossaire .....	8
1    Introduction .....	9
1.1    Contexte et justification .....	9
1.2    Hypothèse .....	10
1.3    Objectifs .....	10
2    Matériel et méthodes .....	11
2.1    Source des données .....	11
2.2    Période d'étude .....	11
2.3    Population .....	11
2.3.1    Critères d'inclusion .....	11
2.3.2    Critères d'exclusion .....	11
2.4    Variables .....	12
2.4.1    Liste des variables .....	12
2.4.2    Choix des regroupements effectués .....	13
2.5    Biais potentiels .....	16
2.5.1    Biais de sélection .....	16
2.5.2    Biais de classement .....	17
2.5.3    Biais de confusion .....	17
2.6    Analyse statistique .....	17
2.6.1    Analyse des données manquantes .....	17
2.6.2    Analyse descriptive .....	19
2.6.3    Création des clusters communaux .....	19
2.6.4    Définition du modèle logistique .....	19
2.6.5    Analyses complémentaires .....	20
3    Résultats .....	21
3.1    Population .....	21
3.2    Description des données .....	21
3.3    Description en fonction du stade au diagnostic .....	22
3.4    Clusters communaux .....	24
3.5    Principaux résultats .....	26
3.5.1    Etude en cas complet .....	26
3.5.2    Etude avec données imputées .....	27
3.6    Résultats complémentaires .....	28

3.6.1	Détail des circonstances de la découverte .....	28
3.6.2	Analyse de la survie .....	29
4	Discussion .....	30
4.1	Résultats clés .....	30
4.2	Limitations .....	31
4.3	Interprétation .....	32
4.4	Validité externe des résultats .....	34
5	Conclusion .....	35
	Annexes .....	36
	Références .....	66

## Glossaire

AIC	Akaike information criterion
Atih	Agence technique de l'information hospitalière
CNIL	Commission nationale de l'informatique et des libertés
DO	Dépistage organisé
DREES	Direction de la recherche, des études, de l'évaluation et des statistiques
ENCR	European network of cancer registries
FDep	Indice français de désavantage social / French deprivation index
Francim	France cancer incidence et mortalité
HAS	Haute autorité de santé
HER2	Human epidermal growth factor receptor 2
INSEE	Institut national de la statistique et des études économiques
INSERM	Institut national de la santé et de la recherche médicale
IRIS	Ilots regroupés pour l'information statistique
RC	Rapport de cotes
RCP	Réunion de concertation pluridisciplinaire
RE	Récepteur œstrogène
RG2C	Registre général des cancers de Corse
RNIPP	Répertoire national d'identification des personnes physiques
RP	Récepteur progestérone
SpF	Santé publique France

# 1 Introduction

## 1.1 Contexte et justification

Le rapport standardisé d'incidence estimé du cancer du sein chez la femme en Corse (vs France métropolitaine) est inférieur à 1 (0,93 IC95[0,89 ; 0,98]), pour une incidence régionale annuelle moyenne estimée de 266 IC95[250 ; 284] sur la période 2007 – 2016 [1]. Le registre général des cancers de Corse (RG2C) a mesuré sur la période 2017 – 2019 une incidence annuelle moyenne des cancers du sein chez la femme de 316, suggérant une potentielle sous-estimation des incidences en Corse [2].

Sur la période 2009 – 2012, 59% des cancers du sein répertoriés ont été diagnostiqués à un stade localisé, 29% à un stade régional et 12% à un stade évolué selon la classification ENCR en France hexagonale [3]. Les données utilisées durant cette étude proviennent des registres des cancers du réseau Francim, qui couvrent environ 20% de la population française. Le registre général des cancers de Corse, opérationnel depuis 2017, est en cours d'agrément par le réseau Francim. Ainsi, ses données ne sont pas encore incluses dans les données du réseau Francim et n'ont pour le moment pas été publiées.

Les facteurs d'un diagnostic évolué peuvent être multiples [4–7]: éloignement d'un centre de dépistage, non-participation au dépistage organisée (DO), délai de prise de rendez-vous, problèmes d'accessibilité pour les personnes non véhiculées liés à une absence de transport public en dehors des zones urbaines, aspects socio-économiques et de défavorisation sociale, aspects psychologiques, etc.

Selon Santé publique France (SpF), le taux de participation au DO du cancer du sein en France est de 50% en 2017 – 2018 [8]. La Corse est la région métropolitaine avec le plus faible taux de participation au DO (35%) sur la même période. Il faut cependant y associer le dépistage individuel, clinique et par imagerie, pratiqué par le médecin traitant, dont l'importance est mal connue [9]. La Corse, en raison de son statut d'île montagne, est doublement pénalisée dans l'accès au DO. Elle est d'une part isolée du reste du territoire continental, et dispose d'autre part d'infrastructures parfois éloignées de certaines populations du fait de sa faible densité en professionnels de santé (comme le montrent des distances isochrones élevées pour se rendre au centre de mammographie le plus proche (cf. Annexe 7, Annexe 9)).

À cela s'ajoutent des facteurs socio-économiques défavorables pour la population insulaire [10]. La Corse est la région la plus pauvre de France métropolitaine avec une forte intensité de la pauvreté (18% de la population vivant sous le seuil de pauvreté vs 14% en moyenne). Les résidents ont également un niveau d'éducation plus faible que la moyenne nationale (hors Mayotte) avec une part de non ou peu diplômés plus élevée (29% vs 26%) et une part de diplômés de l'enseignement supérieur plus faible (27% vs 33%).

À notre connaissance, peu d'études abordent les facteurs expliquant un stade au diagnostic évolué du cancer du sein et aucune sur le territoire Corse. Les études recensées se concentrent sur les facteurs socio-économiques et géographiques[5,7,11] des personnes porteuses d'un cancer du sein mais ces données ne sont pas analysées conjointement avec les données médicales de la tumeur et des circonstances de la découverte.

## 1.2 Hypothèse

L'hypothèse formulée est que la probabilité de diagnostic à un stade évolué du cancer du sein chez la femme en Corse est influencée par des facteurs liés : à la maladie (type morphologique, statut hormonal, ...), au patient (âge, comportement, pratique médicale, ...) et à son environnement (accessibilité potentielle localisée, distance entre le lieu d'habitation et le centre de diagnostic le plus proche, ...).

## 1.3 Objectifs

L'objectif principal de l'étude est l'identification des facteurs associés à un diagnostic à un stade évolué du cancer du sein chez la femme en Corse, incident entre 2017 et 2019.

Afin d'atteindre cet objectif plusieurs étapes sont nécessaires.

- Une description des données en fonction des stades au diagnostic afin de déterminer variable par variable celles potentiellement associées à un stade « évolué » ;
- Un regroupement des communes ayant un profil similaire au regard d'indicateurs médico-sociaux ;
- In fine, la définition d'un modèle logistique permettant de quantifier les effets des variables explicatives sur le stade au diagnostic (évolué ou non).

En complément de l'objectif principal, deux analyses secondaires seront produites :

- Une analyse de sensibilité en sous-groupe en fonction des circonstances de la découverte détaillées en sept catégories afin d'observer d'éventuelles disparités au sein des regroupements des circonstances de la découverte au regard des stades au diagnostic.
- Une analyse de la survie à cinq ans, en fonction du stade au diagnostic, des femmes porteuses d'un cancer du sein afin de déterminer l'impact d'un diagnostic à un stade « évolué » de la maladie en Corse.

## 2 Matériel et méthodes

Pour répondre à l'objectif principal il est proposé de réaliser une étude étiologique transversale rétrospective, incluant l'ensemble des cancers du sein en Corse incidents sur la période 2017 – 2019.

### 2.1 Source des données

Les données nécessaires à l'étude ont été extraites de la base de données du RG2C et enrichie par des données disponibles en opendata, notamment sur le site de l'INSERM pour l'indice de défavorisation sociale, de la DREES pour l'accessibilité potentielle localisée aux médecins généralistes et de l'INSEE pour le fichier des personnes décédés.

S'agissant d'une réutilisation des données du RG2C à des fins de recherche, l'exploitation de la base de données du RG2C est conforme à la réglementation en vigueur et ne nécessite pas de formalités complémentaires auprès de la CNIL ou des personnes concernées car les données publiées sont agrégées ; ce qui ne permet pas l'identification des individus.

### 2.2 Période d'étude

L'extraction des données couvre les cancers du sein incidents de la période comprise entre le 1<sup>er</sup> janvier 2017 et le 31 décembre 2019. Cette période a été définie à la fois afin d'éviter un « effet COVID » qui a potentiellement augmenté les délais de prise en charge à partir du début d'année 2020 en France et donc favorisé un stade au diagnostic évolué, et à la fois afin de pouvoir étudier la mortalité à 5 ans après la date d'incidence du cancer (date de point 1<sup>er</sup> janvier 2025).

L'extraction des données a été réalisée le 15 janvier 2025 et la base de travail figée à cette date.

### 2.3 Population

#### 2.3.1 Critères d'inclusion

La population étudiée est constituée des femmes de plus de 15 ans, porteuses d'un carcinome mammaire invasif incident entre 2017 et 2019. Seules les tumeurs pour lesquels l'ensemble des variables obligatoires selon les règles d'enregistrement nationales et internationales [12,13] sont renseignées par les agents du registre, ont été incluses dans l'étude. Les individus de moins de 15 ans n'ont pas été retenus, afin de ne pas intégrer de pathologies juvéniles qui ne font pas partie de la thématique de recherche.

#### 2.3.2 Critères d'exclusion

Ont été exclues de l'étude :

- les tumeurs sans confirmation histologique ;
- les tumeurs hors carcinomes ;
- les patientes pour lesquelles il n'a pas été possible de déterminer le stade ENCR au diagnostic ;
- les patientes pour lesquelles le résultat de la surexpression d'au moins un des trois marqueurs hormonaux, Récepteur de Progestérone (RP), Récepteur d'Œstrogène (RE) ou Human Epidermal growth factor Receptor 2 (HER2) est absent.

## 2.4 Variables

### 2.4.1 Liste des variables

Se rapportant	Variable	Format	Modalités	Source
à la pathologie	Stade au diagnostic	Discrète	Local limité / Local étendu / Régional / Avancé	RG2C
	Stade au diagnostic en deux catégories	Discrète	Limité / Evolué	RG2C
	Type morphologique de la tumeur	Discrète	Carcinome canalaire / Carcinome lobulaire / Carcinome SAI	RG2C
	Statut hormonal en trois catégories en fonction des marqueurs ER, PR et Her2	Discrète	Basal like / Her2+ / Luminal	RG2C
	Circonstance de la découverte en 3 catégories	Discrète	Patient / Medecin / DO	RG2C
	Circonstance de la découverte en 7 catégories (analyse secondaire)	Discrète	Autopalpation / GG axillaire / Rétraction / Bilan / Dépistage individuel / Examen clinique / DO	RG2C
au patient	Classe d'âge du patient	Discrète	[15 – 40[ / [40 – 50[ / [50 – 75[ / 75 et plus	RG2C
	Statut vital aux dernières nouvelles	Discrète	Oui / Non	RG2C + INSEE
	Code INSEE de la commune de résidence	Discrète	360 codes commune Corse	RG2C
à l'environnement / la commune	Département de résidence	Discrète	Corse-du-Sud / Haute-Corse	RG2C
	Indice de défavorisation sociale (FDep)	Discrète	Quintile d'appartenance	INSERM
	Grille communale de densité	Discrète	Bourg / Rural à habitat dispersé / Rural à habitat très dispersé	INSEE
	Temps de trajet en voiture entre la mairie de la commune de résidence et la mairie d'une commune disposant d'un centre de dépistage la plus proche (distance isochrone)	Discrète	Commune équipée d'un centre / Moins de 15 min / Entre 15 et 30min / Plus de 30 min	RG2C
	Accessibilité potentielle localisée aux médecins généralistes (APL MG)	Discrète	Inférieur à la moyenne régionale / Supérieur à la moyenne régionale	DREES
	Diplôme le plus élevé obtenu par plus de la moitié de la population	Discrète	Inférieur au baccalauréat / Supérieur ou égal au baccalauréat	INSEE

## 2.4.2 Choix des regroupements effectués

### Stade au diagnostic

Le stade au diagnostic est déduit du stade TNM, renseigné selon les règles nationales et internationales en vigueur [12]. Afin de pouvoir confronter les résultats de notre étude aux travaux de SpF [3] la stadification utilisée est quasi-identique, à savoir :

Tableau 1: Stade retenu en fonction de la classification TNM

Stade ENCR	T	N	M
Local limité	1, 2, 3	0	0
Local étendu	4	0	0
Régional	Tous	1,2	0
Avancé	Tous	Tous	1

A la différence de la publication de SpF pour les tumeurs non réséquées la classification TNM clinique a été prise en considération pour définir le stade. L'objectif de l'étude étant de déterminer les facteurs influençant un diagnostic à un stade évolué, le classement en « avancé » de petites tumeurs pour lesquelles il n'a pas été décidé de résection pour des circonstances particulières pour des raisons particulières (mauvais état général, contexte clinique, ...) ne semblait pas pertinent.

Le critère de regroupement permettant de déterminer le stade en deux modalités (Limité ou Evolué) est la diffusion ou non à distance de la tumeur. Les tumeurs limitées à l'organe (local limité et local étendu) ont été recodées en « Limité » et les tumeurs pour lesquels une propagation à distance a été observée (ganglionnaire ou métastatique) ont été recodées en « Evolué ». La diffusion à distance non documentée histologiquement a tout de même été prise en compte à partir du moment où une conviction du clinicien était démontrée (prise en charge thérapeutique adaptée, imagerie, ...).

### Morphologie tumorale

La morphologie tumorale des pathologies éligibles a été recodée en trois catégories en fonction du point d'origine de la tumeur, de son aspect cellulaire [14] et de la disponibilité de l'information. Les catégories conservées sont :

- « Carcinome canalaire infiltrant » pour les carcinomes dont le point de départ se trouve au niveau des canaux galactophores ;
- « Carcinome lobulaire infiltrant » pour les carcinomes dont le point de départ se trouve au niveau des lobules ;
- « Autre carcinome » pour ceux dont le point de départ est incertain ou ne rentre dans aucune des catégories précitées (ex : carcinome mixte).

Le diagnostic clinique et radiologique du carcinome lobulaire infiltrant étant plus difficile [15], les facteurs de risques étant différents [16,17] et le stade au diagnostic pouvant varier selon la morphologie [18], il semblait pertinent de distinguer le groupe morphologique d'appartenance de la tumeur.

### Statut hormonal

Le statut hormonal des tumeurs a été défini en fonction des trois marqueurs suivants : ER, PR et Her2. Cette distinction semblait pertinente au regard des objectifs, la littérature indiquant que la probabilité de découverte de la tumeur à un stade évolué pouvait être influencée par le statut hormonal [14].

Les tumeurs dites « Luminal » regroupent les catégories « Luminal A » et « Luminal B » soit des tumeurs ER+ et / ou PR+. Il n'a pas été possible dans cette étude de réaliser une

distinction entre « Luminal A » et « Luminal B » du fait de l'hétérogénéité de la saisie du marqueur Ki-67 dans le système d'information du RG2C.

Les tumeurs dite « Her2+ » correspondent à des tumeurs dont le statut hormonal est le suivant : ER-, PR- et Her2+.

Les tumeurs dites « Basal like », également appelées triple négatif, correspondent à des tumeurs dont le profil est le suivant : ER-, PR- et Her2-.

### **Circonstances de la découverte**

Les circonstances de la découverte ont été déduites des différents documents à disposition des agents du RG2C. Ces documents comprennent notamment les comptes rendus de réunion de concertation pluridisciplinaire, des comptes rendus d'hospitalisation ou du contexte clinique rappelé dans certains comptes rendus d'anatomopathologie. Les circonstances de la découverte ont été groupées en trois catégories : « Patient », « Médecin » et « DO » (dépistage organisé) et sept sous catégories, précisées ci-dessous.

- La catégorie « Patient » regroupe toutes les circonstances pour lesquelles la patiente a noté une anomalie faisant suspecter un potentiel cancer du sein et qui l'a mené à une consultation qu'elle soit rapide ou retardée. Dans la catégorie « Patient », trois sous-catégories ont été identifiées :
  - o La sous-catégorie « Autopalpation » a été renseignée quand la patiente a réalisé une palpation, a découvert une masse anormale puis a entrepris les démarches afin qu'un diagnostic soit posé.
  - o La sous-catégorie « Rétraction » correspond à l'observation par la patiente d'une anomalie à la surface du sein. Il peut s'agir d'une rétraction du mamelon qui a mené à une consultation puis au diagnostic de la tumeur mais également à une inflammation douloureuse ou d'une façon générale de tout symptôme apparent à la surface du sein ayant mené la patiente à entreprendre des démarches en vue d'un diagnostic.
  - o La sous-catégorie « Ganglion axillaire » correspond à l'inflammation du ganglion axillaire détectée par la patiente. Cette catégorie étant faiblement représentée, les résultats ne seront pas présentés lors de l'analyse du détail des circonstances de la découverte.
- La catégorie « Médecin » regroupe toutes les circonstances pour lesquelles un médecin est à l'origine de la découverte sans informations complémentaires sur une éventuelle suspicion de la part du patient. La catégorie « Médecin » est également subdivisée en trois sous-catégories :
  - o La sous-catégorie « Examen clinique » correspond à un diagnostic posé par le médecin à la suite à une consultation pour un motif n'ayant pas évoqué un cancer à la patiente. Il peut s'agir par exemple d'une consultation pour une altération de l'état général sans motif précis, par la suite les examens réalisés par le médecin ont permis de poser le diagnostic de cancer.
  - o La sous-catégorie « Dépistage individuel » correspond à une mammographie de dépistage prescrit par le médecin et donc sans la double lecture réalisée lors du dépistage organisé. Dans cette catégorie ont été également classées les patientes pour lesquels la notion de dépistage a été évoquée sans précision s'il s'agit d'un dépistage individuel ou organisé.
  - o La sous-catégorie « Bilan » regroupe les découvertes « fortuite » du cancer dans le cadre d'un bilan pour une autre pathologie ou la recherche d'un primitif chez une patiente présentant des métastases.
- La catégorie « DO » a été retenue uniquement si le dépistage organisé a été précisément mentionné dans un document en possession du RG2C. La catégorie « DO » n'a pas de sous-catégorie.

## **Age au diagnostic**

L'âge au diagnostic a été regroupé en quatre catégories définies en fonction des profils pour lesquels des actions de prévention et plus globalement de santé publique peuvent être menées en fonction des résultats de l'étude.

Une première classe d'âge a été définie pour les individus de 15 à 39 ans. Cette classe d'âge correspond aux individus peu touchés par les cancers. Le fait de les isoler et de déterminer d'éventuelles associations avec d'autres co-variables pourra permettre aux acteurs de la santé publique de définir des actions spécifiques à ce jeune public.

La seconde classe d'âge a été définie pour les individus de 40 à 49 ans. Ce sont des individus non éligibles au DO pour lesquels, en fonction des résultats, pourra se poser la question de la nécessité d'un élargissement du DO, d'action de sensibilisation à l'autopalpation, ...

La troisième classe d'âge a été définie pour les individus de 50 à 75 ans. Ce sont des individus éligibles au DO.

Une dernière classe d'âge a été définie pour les individus de plus de 75 ans. Ce sont des individus qui ne sont plus éligibles au DO mais pour lesquels les bonnes pratiques en termes de vigilance au cancer du sein sont restées.

## **Clusters communaux**

Les données socio-démographiques ne sont pas disponibles à l'échelle individuelle au sein du RG2C mais uniquement à l'échelle communale. Afin de s'affranchir des découpages administratives et d'éviter des problèmes de multicolinéarité, dus à une probable corrélation entre les variables socio-démographiques, nous avons choisi, dans le cadre du modèle logistique, de regrouper les communes en clusters de communes similaires.

Ce sont quatre indicateurs qui ont été retenus de par leur disponibilité à l'échelle communale et identifiés comme étant potentiellement en lien avec un stade au diagnostic évolué du cancer du sein dans la littérature [5,7,11] ou par les médecins oncologues du territoire.

## **FDep**

L'indice français de désavantage social (FDep) est un score obtenu à l'échelle communale mesurant le désavantage matériel et social spécifiquement adapté aux études de santé sur la population française [19]. Il est construit à partir du taux de chômage dans la population active de 15 à 64 ans, du taux d'ouvriers dans la population active de 15 à 64 ans, du taux de bacheliers dans la population non scolarisée des plus de 15 ans et du revenu médian déclaré par unité de consommation.

Le FDep a été privilégié à l'european deprivation index (EDI), qui est construit à partir de 9 indicateurs socio-économiques, face à l'hétérogénéité de la répartition des communes insulaires dans les quintiles de l'EDI (cf. Annexe 1).

## **Densité**

La grille communale de densité de l'INSEE en sept niveaux a été utilisée pour définir la taille des communes. La répartition des communes au sein des différents niveaux est hétérogène et un regroupement a donc été effectué. Deux catégories, « Rural à habitat dispersé » et « Rural à habitat très dispersé », ont été conservées et une catégorie, intitulée « Bourg », regroupant les catégories « Centres urbains intermédiaires », « Bourgs ruraux », « Ceintures urbaines » et « Petites villes », a été créée.

## **APL MG**

L'indicateur d'accessibilité potentielle localisée à un médecin généraliste est développé par la DREES permettant de mesurer l'adéquation spatiale entre l'offre et la demande de soins au niveau communal. Il se lit comme un nombre de consultations auprès d'un médecin généraliste libéral accessible à moins de 20 minutes du domicile par habitant standard et par année. En 2018, il est en Corse de 3,2 (vs 3,9 en France) mais avec de fortes disparités communales allant de 0 à 11,1. Les données ont ensuite été recodées en deux catégories : strictement inférieure à la médiane régionale et supérieure à la médiane régionale.

### ***Distance à un centre de mammographie***

La distance au centre de mammographie le plus proche a été calculée à l'aide du package {osrm} [20] du logiciel R. Le temps de trajet calculé est celui, en voiture, de la mairie de la commune de résidence du patient jusqu'à la mairie des communes disposant d'un centre de mammographie. Le plus petit temps de trajet est retenu comme temps de trajet au centre de mammographie le plus proche.

Une géolocalisation exacte, notamment via coordonnées GPS n'était pas possible car hors des centres urbains certaines adresses de résidence sont imprécises (ex : « haut du village » ou « lieu-dit lombarda »).

Pour les « communes éclatées » [21] disposant d'une mairie annexe hors du chef-lieu, quand le code postal de résidence correspondait à la partie de la commune proche de la mairie annexe, alors la mairie annexe a été utilisée comme point de départ du calcul de la distance isochrone.

Les communes ont été regroupées en quatre catégories selon leur distance isochrone la plus proche à un centre de mammographie : « Commune équipée d'un centre », « moins de 15 min », « 15 à 30 min » et « plus de 30 min ».

## **2.5 Biais potentiels**

Des sources de biais potentiels ont été identifiées lors de la rédaction du protocole d'étude. Les biais présentés ci-dessous ne semblent pas « évitables » et doivent donc être connus du lecteur.

### **2.5.1 Biais de sélection**

- Les données du RG2C sont issues d'un solide réseau de partenaires, composé à la fois de l'ensemble des acteurs de la cancérologie insulaire, mais également des principaux centres d'attraction continentaux pour les patientes résidantes dans l'île pour le diagnostic et le traitement de pathologies tumorales. Il n'est cependant pas à exclure que certaines patientes insulaires soient diagnostiquées et suivies exclusivement dans des établissements non-partenaires et ne seraient donc pas détectées dans le système d'information.

Pour limiter ce risque le RG2C réalise régulièrement des requêtes sur la plateforme de l'agence technique de l'information hospitalière (Atih) pour identifier de potentiels centres d'attraction de patientes résidant en Corse. Une convention a été signée avec tous les établissements accueillant plus de 10 patientes en provenance de Corse au cours d'une année (à l'exception de la Clinique Clairval avec laquelle une convention de collaboration a été finalisée et est en cours de signature). Les données de l'Atih sont cependant des données administratives et comptables et non médicales. Il n'est donc pas possible de quantifier le nombre exact d'éventuels cancers incidents non-détectés et présent dans les autres établissements.

- Pour être éligible au RG2C une patiente doit avoir sa résidence principale depuis 5 ans au moins sur le territoire couvert par le RG2C. Il s'agit d'une donnée difficilement vérifiable dans les faits.

En cas de doute de la part de l'équipe du RG2C (parcours de soins incohérent, double résidence, ...) un retour au dossier du patient dans l'établissement de soins assurant la prise en charge de la pathologie est effectué de façon systématique.

### 2.5.2 *Biais de classement*

Le codage des tumeurs se fait selon des règles strictes, fixés par l'Organisation Mondiale de la Santé [13], et l'Union for International Cancer Control [12]. Cependant il n'est pas à exclure que des erreurs humaines soient à l'origine d'erreurs de saisies, notamment dans le stade au diagnostic.

Afin de limiter ces erreurs, le personnel du RG2C a suivi la formation « Techniques de surveillance épidémiologique, d'enregistrement et de codage des cancers » organisées par Francim et SpF.

### 2.5.3 *Biais de confusion*

L'état psychique de la patiente ou d'éventuels problèmes sociaux non connus (consommation de stupéfiant, alcoolisme chronique, aspect psychologique, ...) peuvent influer le stade au diagnostic, la prise en charge et la mortalité des patientes porteuses d'un cancer du sein [22-24]. Ces éléments ne sont pas collectés en routine au RG2C et sont par conséquent non disponibles.

La non prise en compte de ces facteurs fera partie des limites de cette étude et sera intégrée dans la discussion des résultats.

## 2.6 Analyse statistique

L'analyse statistique a été réalisée avec le logiciel R, version 4.4.2,. L'intégralité des informations de la session R utilisée pour produire les analyses est disponible en Annexe 2.

### 2.6.1 *Analyse des données manquantes*

L'analyse des données manquantes a pour objectif de définir la méthode de gestion des données manquantes.

#### Dénombrement

Le jeu de données contenant l'ensemble des tumeurs éligibles est composé de 902 tumeurs. Il contient trois types des données manquantes :

- Pour 8 tumeurs, soit moins de 1% des tumeurs, il n'a pas été possible de déterminer avec certitude le stade au diagnostic de la tumeur. Ces tumeurs ont été exclues de l'analyse.
- Pour 10 tumeurs, soit 1% des tumeurs, le statut hormonal est non connu (au moins un des marqueurs manquant). Ces tumeurs ont été exclues de l'analyse.
- Pour 70 tumeurs, soit 8% des tumeurs, les circonstances de la découverte n'ont pu être déterminées précisément ou n'étaient pas indiquées dans les documents à disposition du RG2C. Ce nombre étant non négligeable il a été décidé de définir une méthode de gestion des données manquantes

Le jeu de données utilisé pour l'étude, après l'inclusion des tumeurs éligible et les exclusions dues aux valeurs absentes, comporte 884 tumeurs dont 70 pour lesquelles les circonstances de la découverte ne sont pas connues.

## Choix de la méthode de gestion des données manquantes

Lors de l'analyse des données manquantes une surreprésentation d'individus de Haute-Corse et avec un stade au diagnostic limité a été constaté. Après concertation avec les oncologues du RG2C les hypothèses suivantes ont été formulées :

- Lorsqu'un retour au dossier n'a pas été nécessaire pour l'enregistrement des variables obligatoires dans le cadre d'un registre des cancers, les circonstances de la découverte ont été collectées dans la plupart des cas à partir des éléments disponibles dans le compte rendu de réunion de concertation pluridisciplinaire (RCP). Or, les cancers du sein à un stade limité, dont la prise en charge est moins complexe, ne sont pas tous discutés en RCP et donc il a été pour certains d'entre eux impossible de renseigner les circonstances de la découverte.
- La surreprésentation des individus de Haute-Corse pour lesquels les circonstances de la découverte sont inconnues est probablement due à la différence d'adressage en bord à bord en cas de prise en charge sur le continent des résidents Corse. Les individus résidant en Corse-du-Sud ont plutôt tendance à être adressés dans des établissements marseillais alors que les individus résidant en Haute-Corse ont tendance à être adressés dans des établissements niçois. L'impossibilité, pour des raisons techniques, d'extraire l'intégralité des comptes rendus de RCP du Centre Antoine Lacassagne (Nice), explique probablement cette surreprésentation des individus résidant en Haute-Corse.

La probabilité qu'une donnée soit manquante ne semblait donc pas dépendre de la valeur de cette donnée mais des valeurs d'autres données (ici : le stade au diagnostic et le département de résidence). Nous faisons donc l'hypothèse d'un mécanisme de données manquantes « missing at random » ce qui légitime l'imputation multiple pour estimer ces données.

## Imputation multiple

L'imputation multiple a été réalisée à l'aide du package `{mice}` [25] de R. Le package `{mice}` permet de créer à partir du jeu de données incomplet plusieurs jeux de données contenant des données imputées à l'aide de processus intégrant une part d'aleatoire, de réaliser des analyses sur l'ensemble des jeux de données puis de regrouper les résultats (en moyennant) afin d'obtenir une base de données imputées (cf. Figure 1).

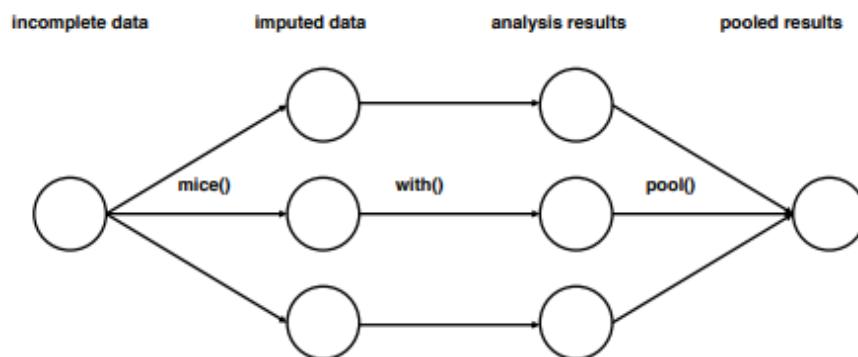


Figure 1 : Principales étapes de l'imputation multiple avec le package mice [Ref package p.5/67]

Le nombre d'imputations proposé en utilisant la méthode décrit par Graham [26], qui est de 2, semblait faible et le nombre de tumeurs incluses dans l'étude permet un plus grand nombre d'imputations. Nous avons donc opté pour une analyse des résultats des données imputées en augmentant progressivement le nombre d'imputation jusqu'à stabilisation des proportions des modalités imputées. Pour cela il a été testé différents nombres d'imputations. Pour notre étude la stabilisation des résultats s'observait à partir de  $m = 10$  imputations (cf. Annexe 3).

### 2.6.2 Analyse descriptive

Afin de déterminer quelles variables présentent une association brute significative avec un stade au diagnostic évolué, des tests de comparaison du Khi<sup>2</sup> ont systématiquement été effectués.

L'analyse descriptive a été effectuée en cas complet (814 tumeurs) et également sur le jeu de données imputées (884 tumeurs).

### 2.6.3 Création des clusters communaux

La création des profils de communes a été réalisée par une méthode d'apprentissage non supervisée. Les variables disponibles permettant de créer des clusters communaux sont le Fdep en quintile de la population, la distance isochrone à un centre de dépistage en quatre catégories, l'APL en deux catégories et le type de commune en trois catégories.

Une classification hiérarchique ascendante a été réalisée pour la création des profils communes selon la méthode de Ward [27]. Le nombre de cluster a été déterminé à l'aide du dendrogramme, en fonction de la dissimilarité inter-classes, de la cohérence et de l'interprétabilité des classes créées.

### 2.6.4 Définition du modèle logistique

Afin de sélectionner les variables à intégrer au modèle logistique et les interactions potentielles entre elles, une matrice de corrélation a été réalisée pour évaluer leur indépendance 2 à 2. Les variables suivantes ont été testées : la classe d'âge, le type cellulaire, le statut hormonal, les circonstances de la découverte et le cluster communal. Les couples de variables dont la valeur du V de Cramer est inférieure à 0,20 ont été intégrées dans le modèle car considérées comme indépendantes. Pour les couples de variables dont la valeur du V de Cramer est comprise entre [0,20 ; 0,50] l'interaction est testée dans le modèle afin de représenter cette association. Pour les couples de variables donc la valeur du V de Cramer est supérieure à 0,50, il a été décidé d'exclure l'une ou l'autre selon leur importance et leur dépendance à d'autres variables.

Un modèle logistique complet : Stade évolué de la maladie au diagnostic (oui/non) en fonction de l'ensemble des variables disponibles et leur interaction calculée a été réalisé. Un second modèle, plus économique en variable, basé sur le critère d'AIC est également proposé pour chacun des jeux de données (cas complet et avec données imputées). La formule mathématique du modèle logistique est présentée en Annexe 4.

L'association globale de chaque variable avec le Stade au diagnostic a été testée via un test de rapport de vraisemblance. Les variables considérées comme significativement associées au stade au diagnostic sont celles dont la p-value du test est inférieure au seuil de 0,05. Pour le jeu de données imputées, pour chaque variable a été réalisé un test d'association : le test D3 [28,29]. Ce test spécifique aux jeux de données imputées compare les modèles sur l'ensemble des jeux de données imputées, de façon comparable à une ANOVA, permettant d'obtenir une statistique F et ainsi de tester la significativité de l'association d'une variable du modèle sur le stade au diagnostic.

Le rapport de cotes (RC) ainsi que l'intervalle de confiance à 95% pour celui-ci ont été calculés, à partir des coefficients de régression, pour chaque modalité par rapport à la modalité de référence sur la probabilité d'avoir un stade au diagnostic évolué.

## 2.6.5 Analyses complémentaires

### Détails des circonstances de la découverte

Les détails des circonstances de la découverte ont été comparés en complément de l'objectif principal afin de déterminer d'éventuelles différences au sein des trois circonstances de la découverte définies pour répondre à l'objectif principal.

La taille des effectifs étant faible au sein des sous-catégories aucune modélisation n'a été réalisée et leur analyse descriptive permet uniquement d'alimenter la discussion.

### Analyse de la survie

Le statut vital de chaque patient a été renseigné à partir de deux sources de données : les informations collectées en routine et le croisement avec le répertoire national d'identification des personnes physiques (RNIPP). Le croisement a été effectué sur les variables suivantes : nom (ou nom de naissance), prénom, sexe, date de naissance et commune de naissance (si disponible). La date de fin de suivi correspond soit au décès (le cas échéant) ou à la date de dernières nouvelles. Les données ont été censurées cinq ans après la date d'incidence.

Le RG2C ne disposant pas des causes de décès, l'analyse de la survie est donc une analyse de la survie globale à cinq ans après le diagnostic. Elle est stratifiée en deux groupes définis selon le stade au diagnostic.

Afin de comparer les courbes de survie des deux groupes un test du log-rank a été réalisé et la p-value calculée.

Une analyse de la mortalité, basée sur le modèle décrit ci-dessus, pour l'ensemble des variables pour lesquelles une association brute avec le stade au diagnostic a été mise en évidence est disponible en annexe.

### 3 Résultats

#### 3.1 Population

La base de données du RG2C contenait, au 15 janvier 2025, 2096 tumeurs incidentes et validées du sein chez des patientes résidentes en Corse (cf. Annexe 1). Au total, ce sont 1194 tumeurs qui ne sont pas éligibles pour cette étude, dont 1063 qui ont une date d'incidence après 2019, 114 tumeurs in-situ, 6 sans confirmation histologique et 11 ayant des morphologies non éligibles : soit 902 tumeurs éligibles.

Sur les 902 tumeurs éligibles, le stade au diagnostic n'a pas pu être déterminé par les agents du RG2C avec certitude pour 8 tumeurs soit moins d'1% des cas. La recherche des stades au diagnostic de ces 8 cas aurait nécessité des efforts disproportionnés pour les agents du RG2C il a donc été décidé de supprimer ces 8 cas de l'étude.

Sur les 894 tumeurs éligibles et avec un stade au diagnostic connu, pour 10 d'entre elles, soit 1%, le statut hormonal n'était pas disponible (illisible sur compte rendu d'anatomopathologie ou non mesuré). Pour 4 la présence du récepteur de progestérone n'a pas pu être mesuré et pour 2 la présence du récepteur d'oestrogène. Parmi les 10 comptes rendus pour lesquels le statut hormonal n'est pas disponible, la recherche du nombre de récepteurs de la protéine Her2 n'a pas été réalisée pour 9. La part de tumeurs pour lesquels cette information est manquante étant faible il a été décidé d'exclure ces tumeurs de l'étude.

Sur les 884 tumeurs éligibles et non exclues, pour 70, soit 8%, d'entre elles les circonstances de la découverte n'étaient pas connues. Une méthode d'imputation multiple a été retenue (cf. 2.6.1).

Les résultats présentés dans la suite du document ont donc été calculés à partir d'un jeu de données contenant 814 individus pour les analyses en cas complet et sur un jeu de données contenant 884 individus pour le jeu avec données imputées.

#### 3.2 Description des données

La répartition des cancers du sein selon le stade ENCR en quatre catégories est hétérogène. La majorité des cancers du sein, 587/884 soit 66% (cf. Annexe 6), ont un stade au diagnostic « local limité ». Cette proportion est similaire dans le jeu de données contenant les données imputées et celui contenant uniquement les données complètes (cf. Annexe 8). Après regroupement des stades « local limité » et « local étendu », ce sont 597/884 tumeurs qui ont un stade limité, soit 68% dans le jeu de données contenant des valeurs absentes vs 548/814, soit 67% dans le jeu de données « cas complet ».

L'âge moyen des patientes, identique à l'âge médian dans les deux jeux de données, est de 65 ans. La patiente la plus jeune à 26 ans et la plus âgée 98 ans. La courbe de densité de répartition des patientes selon l'âge a une forme unimodale symétrique, d'allure gaussienne (cf. Annexe 7 et Annexe 9).

Pour plus de la moitié des tumeurs (56%) le médecin est à l'origine de la découverte de la tumeur. Parmi les 458/814 tumeurs dont le médecin est à l'origine de la découverte, dans 287/458 cas (63%) il s'agit d'une mammographie prescrite par le médecin, pour 117/458 cas (26%) la découverte a été réalisée lors d'un examen clinique et pour les 54/458 cas restant la découverte a été réalisée lors d'un bilan. Le patient est à l'origine de la découverte de sa tumeur dans 206/814 cas (25%). Le cas de figure le plus récurrent est une autopalpation (153/206 soit 74%), parfois des symptômes cutanés type rétractation mamelonnaire ou syndrome inflammatoire (49/206 soit 24%) et à la marge une inflammation du ganglion axillaire (4/206 soit 2%). Le dépistage organisé est à l'origine de la découverte de 150/814 tumeurs sur la période, soit 18% des tumeurs.

Dans les deux jeux de données la grande majorité des tumeurs (85% en cas complet et 86% pour les données imputées) ont un statut hormonal de type « Luminal ». Les tumeurs de type « Basal like » sont le deuxième type de tumeurs le plus fréquent (respectivement 8,7% et 8,3%). Les tumeurs Her2 surexprimé ne représentent que 6% des tumeurs dans les deux jeux de données.

La répartition des tumeurs selon le département de résidence est similaire tant dans l'analyse en cas complet que dans le jeu de données imputées. La Haute-Corse, département le plus peuplé de l'île, recense un nombre de tumeurs plus important que la Corse-du-Sud (51% vs 49% en cas complet et 52% vs 48% pour le jeu de données avec imputation).

Le temps de trajet moyen jusqu'au centre de mammographie le plus proche est de 9 minutes et la médiane de 14 minutes dans les deux jeux de données. Cependant, on observe une forte concentration des patientes avec des temps de trajet très cour (45% de la population vit dans une commune disposant d'un centre de mammographie). A contrario, moins de 20% de la population vit dans une commune à plus de 30min d'un centre de mammographie, cette distance isochrone pouvant aller jusqu'à 86min.

Les patientes porteuses d'un cancer du sein résident majoritairement dans des communes du deuxième décile du Fdep (38% dans l'analyse en cas complet et 37% dans le jeu de données avec données imputées). Les patientes résidentes dans des communes du premier décile sont les plus rares (7% dans les deux jeux de données).

Dans les deux jeux de données 67% résident dans des communes dans lesquelles plus de la moitié des résidents ont un diplôme supérieur ou égal au baccalauréat et dans les deux jeux de données 74% résident dans des communes dont l'APL est supérieur à la moyenne.

### 3.3 Description en fonction du stade au diagnostic

Environ 75 % des cancers du sein inclus dans l'étude sont des carcinomes canalaire infiltrant et 15% des carcinomes lobulaires infiltrants. Cette proportion ne varie pas en fonction du stade au diagnostic en deux catégories (cf. Annexe 10). Il n'a pas non plus été constaté d'association brute entre l'année de diagnostic, le statut hormonal, les autres variables socio-démographiques et le stade au diagnostic en deux catégories.

Il a cependant été constaté une très forte association brute dans les deux jeux de données (p-value << 0,01) entre les circonstances de la découverte et le stade au diagnostic. Parmi les personnes ayant un stade limité au diagnostic, la proportion de personnes dont le diagnostic a été posé lors d'un dépistage organisé est plus importante que pour les personnes à un stade évolué (23% vs 9%). A contrario, la proportion à un stade limité ayant eu des signes d'appels leur faisant suspecter un cancer du sein est plus faible (19% vs 38%). Le médecin reste cependant la personne ayant amené le plus souvent au diagnostic que ce soit pour les stades limités ou pour les stades évolués (58% et 53%).

Il existe également une association brute entre la classe d'âge des individus et le stade au diagnostic (p-value < 0,05). Les individus ayant un stade au diagnostic limité sont plus souvent dans la classe d'âge [50-75[ans, pour 361/884 soit 60% vs 151/884 soit 53% pour les individus ayant un stade au diagnostic évolué. A contrario, les individus ayant un stade limité ont moins souvent 50 ans ou plus : 2,8% ont entre 15 et 40 ans (vs 5,6% pour les stades évolués) et 12% ont entre 40 et 50 ans (vs 16% pour les stades évolués).

Une association brute entre le département de résidence et le stade au diagnostic (p-value < 0,01) a également été mise en évidence avec une surreprésentation d'individus résidant en Corse-du-Sud ayant un stade au diagnostic évolué (55% vs 45% pour la Haute-Corse). A contrario, 44% des individus ayant un stade au diagnostic limité sont résidents en Corse-du-Sud (vs 56% de Haute-Corse).

Tableau 2 : Description du jeu de données en fonction du stade au diagnostic en deux catégories, jeu de données avant imputation (N = 884)

Caractéristiques	limité N = 597 <sup>1</sup>	étendu N = 287 <sup>1</sup>	p-value <sup>2</sup>
<b>Morphologie</b>			<b>&gt;0.9</b>
Carcinome canalaire	449 / 597 (75%)	214 / 287 (75%)	
Carcinome lobulaire	92 / 597 (15%)	46 / 287 (16%)	
Carcinome SAI et autres	56 / 597 (9.4%)	27 / 287 (9.4%)	
<b>Classe d'âge</b>			<b>0.033</b>
[15-40[	17 / 597 (2.8%)	16 / 287 (5.6%)	
[40-50[	70 / 597 (12%)	46 / 287 (16%)	
[50-75[	361 / 597 (60%)	151 / 287 (53%)	
75 et +	149 / 597 (25%)	74 / 287 (26%)	
<b>Année de diagnostic</b>			<b>0.2</b>
2017	209 / 597 (35%)	86 / 287 (30%)	
2018	196 / 597 (33%)	93 / 287 (32%)	
2019	192 / 597 (32%)	108 / 287 (38%)	
<b>Circonstance de la découverte</b>			<b>&lt;0.001</b>
DO	126 / 548 (23%)	24 / 266 (9.0%)	
Médecin	318 / 548 (58%)	140 / 266 (53%)	
Patient	104 / 548 (19%)	102 / 266 (38%)	
<b>Statut hormonal</b>			<b>0.4</b>
Basal like	48 / 597 (8.0%)	25 / 287 (8.7%)	
Her2 surexprimé	32 / 597 (5.4%)	22 / 287 (7.7%)	
Luminal	517 / 597 (87%)	240 / 287 (84%)	
<b>Département</b>			<b>0.002</b>
Corse-du-Sud	263 / 597 (44%)	158 / 287 (55%)	
Haute-Corse	334 / 597 (56%)	129 / 287 (45%)	
<b>FDep</b>			<b>0.8</b>
1 (Très favorisé)	41 / 597 (6.9%)	22 / 287 (7.7%)	
2	228 / 597 (38%)	98 / 287 (34%)	
3	140 / 597 (23%)	71 / 287 (25%)	
4	120 / 597 (20%)	63 / 287 (22%)	
5 (Très défavorisé)	68 / 597 (11%)	33 / 287 (11%)	
<b>APL</b>			<b>0.3</b>
Inférieur à la moyenne	162 / 597 (27%)	69 / 287 (24%)	
Supérieur à la moyenne	435 / 597 (73%)	218 / 287 (76%)	
<b>Baccalauréat</b>			<b>0.9</b>
Non	195 / 597 (33%)	95 / 287 (33%)	
Oui	402 / 597 (67%)	192 / 287 (67%)	

Caractéristiques	limité N = 597 <sup>1</sup>	étendu N = 287 <sup>1</sup>	p-value <sup>2</sup>
<b>Temps de trajet</b>			<b>0.5</b>
Commune avec Mammo	258 / 597 (43%)	136 / 287 (47%)	
Moins de 15min	102 / 597 (17%)	43 / 287 (15%)	
Entre 15 et 30min	120 / 597 (20%)	60 / 287 (21%)	
Plus de 30min	117 / 597 (20%)	48 / 287 (17%)	

<sup>1</sup>n / N (%)

<sup>2</sup>Pearson's Chi-squared test

### 3.4 Clusters communaux

La Corse compte 360 communes dont 124 en Corse-du-Sud et 236 en Haute-Corse. Les communes ont été réparties en cinq classes en fonction du dendrogramme (cf. Annexe 13) généré lors de la classification hiérarchique ascendante et de la cohérence des classes créées.

La première classe créée est composée de 42 communes. Elle se caractérise par un niveau de déprivation sociale faible, avec 31% des communes appartenant au premier quintile de déprivation sociale et 60% au second. Toutes les communes de la classe communale se situent à moins de 30 minutes d'un centre de mammographie et 60% à moins de 15 minutes. 33% des communes de la classe ont été groupées dans la catégorie communale « Bourg » et 64 % dans la catégorie « Rural dispersé ». Dans la suite de l'étude cette classe communale, au vu des caractéristiques des communes la composant, sera dénommée « Urbain, favorisé ».

La seconde classe créée est composée de 87 communes. Elle se caractérise également par un niveau de déprivation sociale faible avec l'ensemble des communes appartenant aux deux premiers quintiles de déprivation dont 33% au premier. C'est la classe contenant le plus de communes avec un APL inférieur à la moyenne régionale (85%). Toutes les communes se situent à plus de 15 minutes d'une commune avec mammographie dont 79% à plus de 30 minutes. Les communes sont essentiellement réparties entre communes « Rural dispersé », pour 38% d'entre elles et 59% sont des communes « Rural très dispersé ». Dans la suite de l'étude cette classe communale, au vu des caractéristiques des communes la composant sera dénommée « Rural, favorisé ».

La troisième classe est composée de 63 communes. Elle se caractérise par un niveau de déprivation sociale plus important que les deux classes précédentes avec 70% des communes de la classe appartenant au quatrième quintile de déprivation sociale, les 30% restant se répartissant entre le troisième et le cinquième quintile. Près de la moitié (48%) des communes de la classe ont une APL supérieur à la moyenne régionale, c'est la classe contenant en proportion le plus de commune avec un APL supérieur à la moyenne régionale. Pour 71% des communes de la classe le centre de mammographie le plus proche se situe entre 15 minutes et 30 minutes de la commune. Il s'agit d'une classe contenant des communes plutôt peuplées car 37% des communes de la classe sont de type « Bourg » et 44% de type « Rural dispersé ». Dans la suite de l'étude cette classe communale, au vu des caractéristiques des communes la composant sera dénommée « Urbain, défavorisé ».

La quatrième classe est composée de 89 communes. Elle se caractérise par un niveau de déprivation sociale intermédiaire, 87% des communes appartiennent au troisième quintile. Les communes sont plutôt éloignées du centre de mammographie le plus proche avec 24% des communes se situant entre 15 et 30 minutes d'un centre de mammographie et 76% à plus de 30 minutes. Les communes de la classe sont toutes rurales dont 60% « Rural dispersé »

et 40% « Rural très dispersé ». Dans la suite de l'étude cette classe communale, au vu des caractéristiques des communes la composant, sera dénommée « Rural, intermédiaire ».

La cinquième et dernière classe créée est composée de 79 communes. Elle se caractérise par l'appartenance de communes défavorisées, 48% des communes dans le quatrième quintile de déprivation sociale et 52% dans le cinquième. Les communes la constituant sont éloignées d'un centre de mammographie (90% à plus de 30 minutes et les 10% restant à plus de 15 minutes). Les communes la composant sont pour 51% d'entre elles de type « Rural dispersé » et 49% « Rural très dispersé ». Dans la suite de l'étude cette classe communale, au vu des caractéristiques des communes la composant, sera dénommée « Rural, défavorisé ».

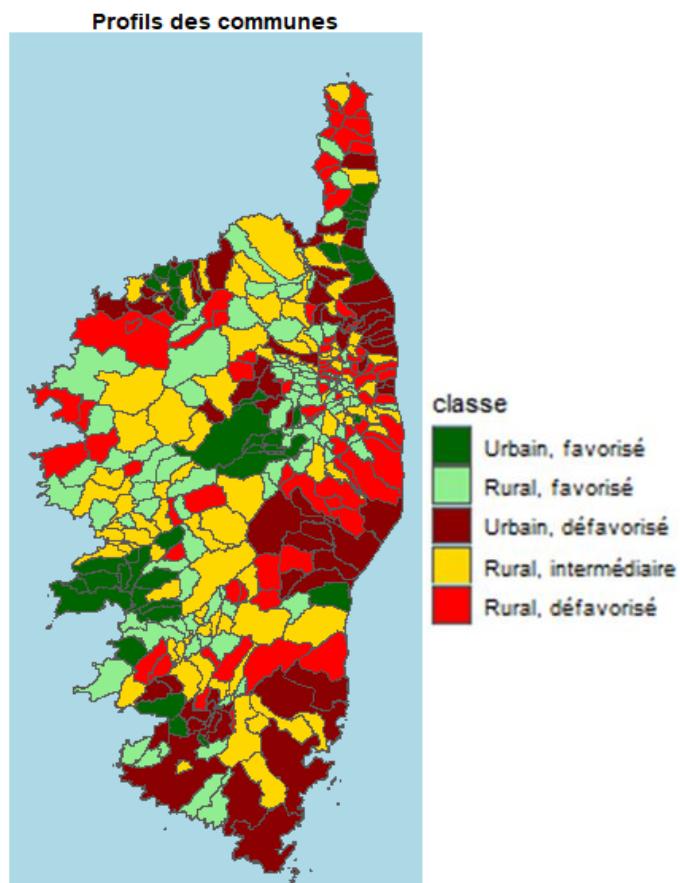


Figure 2 : Cartographie des communes Corse selon leur classe communale

La cartographie ci-dessus représente les communes selon leur classe d'appartenance. Nous repérons les zones urbaines favorisées à proximité des communes plus importantes : la région Ajaccienne descendant jusqu'à Pietrosella, la région Bastiaise, les communes autour de Corte, L'Ile-Rousse et Olmetto / Viggianello. Ventiseri est également associé à cette classe, probablement due à la présence de la base militaire et donc la présence de nombreux militaires au pouvoir d'achat élevé par rapport au reste de la région. Les zones urbaines défavorisées sont également identifiables notamment : la commune de Bastia et les communes de Borgo a Penta-di-Casinca, la zone autour d'Aléria / Ghisonaccia, Porto-Vecchio / Bonifacio, Sartène / Propriano. D'une façon générale, il semble se dégager des zones rurales plus fragiles du côté oriental de l'île et dans le cap Corse et à contrario des zones rurales plus favorisées autour du très grand Ajaccio et du Cortenais. La liste des libellés des communes par classe communale est disponible en Annexe 15.

### 3.5 Principaux résultats

#### 3.5.1 Etude en cas complet

##### Modèle complet

Afin de déterminer les éventuelles interactions entre les variables à inclure dans le modèle, une matrice de corrélation a été établie (cf. Annexe 16). Aucune corrélation inter-variables n'excède 0,20. Aucune interaction n'a donc été prise en compte et l'ensemble des variables a été inclus dans le modèle.

Une analyse de la déviance a été réalisée pour l'ensemble des variables présentes dans le modèle (cf. Annexe 17). La seule variable dont l'association est significative au risque alpha = 5% est « Circonstance de la découverte » (p-value << 0,01).

Les modalités de la variable « Circonstance de la découverte » sont les seules dont le RC est strictement différent de 1 (cf. Annexe 18). Toutes choses égales par ailleurs, la cote de découverte d'un cancer à un stade évolué quand le médecin est à l'origine de la découverte est 2,25 IC95[1,40 ; 3,73] fois plus important que pour un cancer découvert lors d'un dépistage organisé. La cote de découverte à un stade évolué augmente d'avantage lorsque le patient est à l'origine de la découverte avec un RC de 5,01 IC95[2,97 ; 8,69]. Il est à noter que le risque augmente de 30% IC95[0,92 ; 1,83] pour les individus résidant dans une commune appartenant à la classe communale « Urbain, défavorisé » et de 68% IC95[0,93 ; 3,01] pour les individus résidante dans une commune appartenant à la classe communale « Rural, intermédiaire » par rapport à la classe communale de référence, ici « Urbain, favorisé ». Cette augmentation de risque n'est cependant pas statistiquement significative au risque alpha = 5% dans notre étude.

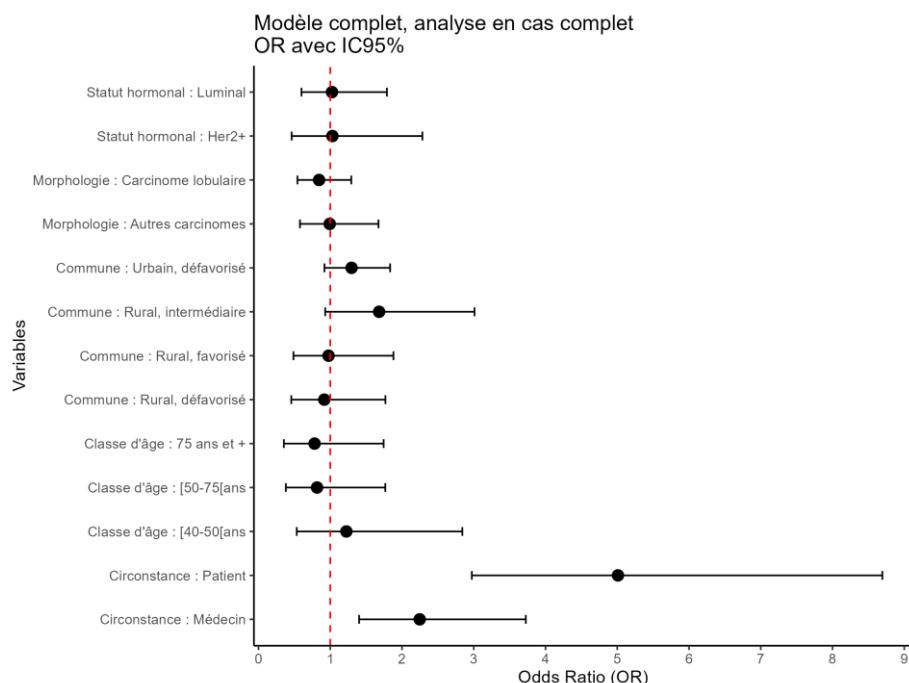


Figure 3 : Diagramme en barre des rapports de cotes et leur intervalle de confiance des différentes modalités des variables intégrées dans le modèle complet dans le cadre de l'analyse en cas complet

##### Modèle restreint

Afin d'éviter un éventuel surajustement et de créer un modèle plus simple, une méthode de sélection descendante des variables à partir du modèle complet a été réalisé. La sélection s'est faite en se basant sur l'Akaike Information Criterion (AIC).

Le modèle restreint finalement retenu ne contient que les circonstances de la découverte, faisant passer la valeur de l'AIC de 1000 pour le modèle complet à 987.

Les RC obtenus avec le modèle restreint sont relativement similaires à ceux obtenu à partir du modèle complet. L'augmentation du risque de découverte d'un cancer à un stade évolué lorsque le médecin est à l'origine de la découverte par rapport au dépistage organisé est de 2,31 IC95[1,45 ; 3,81]. L'augmentation du risque de découverte d'un cancer à un stade évolué lorsque le patient est à l'origine de la découverte par rapport au dépistage organisé est de 5,15 IC95[3,12 ; 8,77].

### 3.5.2 Etude avec données imputées

#### Modèle complet

Afin de déterminer les éventuelles interactions entre les variables à inclure dans le modèle, une moyenne des matrices de corrélation a été établie (cf. Annexe 16). Aucune corrélation inter-variables n'excède 0,20. Aucune interaction n'a donc été prise en compte et l'ensemble des variables a été inclus dans le modèle.

Une analyse de la déviance a été réalisée pour l'ensemble des variables présentes dans le modèle (cf. Annexe 19). La seule variable dont l'association est significative au risque alpha = 5% est « Circonstance de la découverte » (p-value << 0,01).

Dans le cas de l'analyse avec le jeu de données imputées les estimations des RC sont similaires à ceux de l'analyse en cas complets. Toutes choses égales par ailleurs, le risque de découverte d'un cancer à un stade évolué quand le médecin est à l'origine de la découverte est 2,06 IC95[1,27 ; 3,34] fois plus important que lors d'un diagnostic via le dépistage organisé. Le risque de découverte d'un cancer à un stade évolué augmente d'avantage lorsque le patient est à l'origine de la découverte avec un RC égal à 4,18 IC95[2,43 ; 7,18]. Tout comme pour l'analyse en cas complets, les résidentes de communes appartenant à la classe commune « Urbain, défavorisé » ont une augmentation du risque de +16% IC95[0,84 ; 1,62] et ceux de la classe communale « Rural, intermédiaire » de +57% IC95[0,90 ; 2,72]. Cette augmentation du risque n'est cependant pas significative au seuil alpha = 5% et les intervalles de confiance sont plus centrés autour de la valeur 1 que dans l'analyse en cas complets.

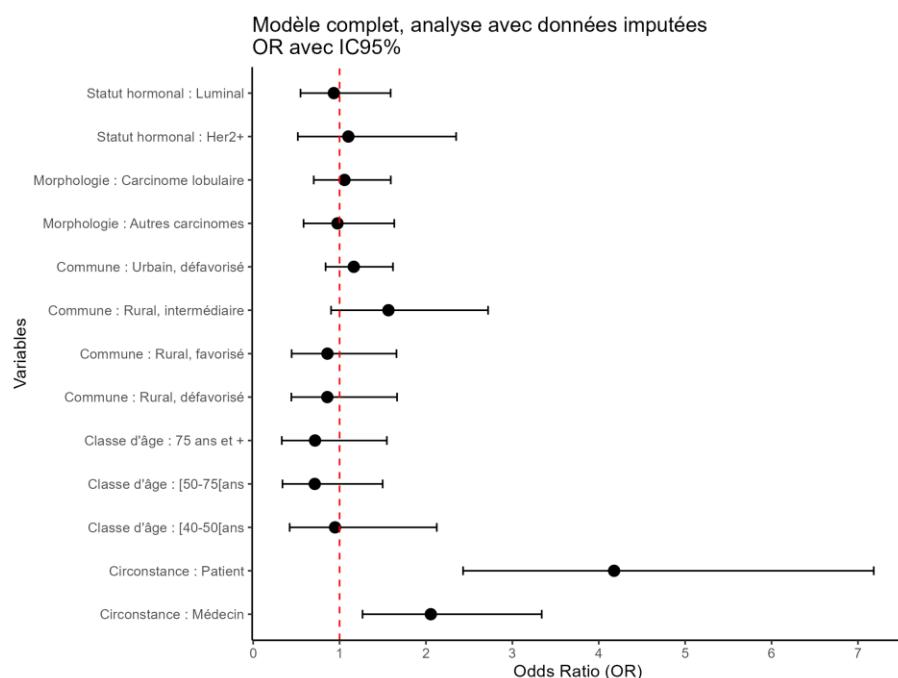


Figure 4 : Diagramme en barre des rapports de cotes et leur intervalle de confiance des différentes modalités des variables intégrées dans le modèle complet dans le cadre de l'analyse des données imputées

## Modèle restreint

Afin d'éviter un éventuel surajustement et de créer un modèle plus simple, une méthode de sélection descendante des variables à partir du modèle complet a été réalisée. La sélection s'est faite en se basant sur l'Akaike Information Criterion (AIC).

Le modèle restreint finalement retenu ne contient que les circonstances de la découverte. Les estimations moyennes des RC et intervalles de confiance obtenus avec le modèle restreint sont relativement similaires à ceux obtenus à partir du modèle complet. L'augmentation du risque de découverte d'un cancer à un stade évolué lorsque le médecin est à l'origine de la découverte par rapport au dépistage organisé est de 2,14 IC95[1,33 ; 3,44]. L'augmentation du risque de découverte d'un cancer à un stade évolué lorsque le patient est à l'origine de la découverte par rapport au dépistage organisé est de 4,46 IC95[2,66 ; 7,50].

## 3.6 Résultats complémentaires

### 3.6.1 Détail des circonstances de la découverte

Les données suivantes sont exclusivement issues de l'analyse en cas complet afin de décrire des données observées et non imputées.

Parmi les personnes dont le diagnostic a été posé lors du dépistage organisé, modalité de référence dans le modèle logistique, 84% (n = 126) ont un stade au diagnostic « limité » (cf. Annexe 21). Parmi les 16% (n = 24) dont le stade au diagnostic est « évolué » une seule femme a eu un diagnostic à un stade métastatique soit 4% des stades au diagnostic « évolué » et moins de 1% de l'ensemble des tumeurs.

Parmi les personnes dont le diagnostic a été posé par le médecin on observe des disparités de stade en fonction des circonstances détaillées. Les personnes dont la découverte a été réalisée lors d'un bilan sont proportionnellement celles dont le stade au diagnostic est le plus souvent « évolué » dans la catégorie « Médecin » (39% (n = 21) de stade « évolué », dont 57% (n = 12) à un stade métastatique). A contrario, ceux dont le diagnostic a été posé lors d'un dépistage individuel ont le moins souvent un stade un diagnostic « évolué » (28% (n = 81), dont 16% (n = 13) à un stade métastatique). Les personnes dont le diagnostic a été posé lors d'un examen clinique ont un stade « évolué » dans 32% (n = 38) des cas. La classe d'âge « [50 ; 75[ ans » est surreprésenté dans la souspopulation de personnes dont le cancer est découvert suite à un dépistage individuel. A contrario, la classe d'âge « 75 ans et + » est surreprésenté chez les individus dont le diagnostic a été posé suite à un examen clinique ou un bilan.

Parmi les patientes qui ont détecté des symptômes ayant conduit au diagnostic, une majorité a découvert une grosseur à la suite d'une autopalpation (77% ; n = 148). La classe d'âge [15 ; 40[ ans est surreprésentée parmi les personnes ayant découvert leur tumeur à la suite d'une autopalpation (9% vs 4% dans l'ensemble de l'échantillon). Tout comme pour le dépistage organisé ou le dépistage individuel il s'agit plutôt de patientes résidante dans des communes de type « Urbain, favorisé » pour environ 40% d'entre elles (vs 30% pour les autres circonstances détaillées de la découverte). La circonstance de la découverte « Autopalpation » est la deuxième plus défavorable avec 49% (n = 72) des cancers découverts à un stade « évolué ». La circonstance de la découverte « Rétraction mamelonnaire » est la plus défavorable avec 58% (n = 26) des tumeurs découvertes à un stade « évolué », dont 35% (n = 9) à un stade métastatique. Les circonstances de la découverte « Rétraction mamelonnaire » sont celle avec, en proportion, le plus de patientes de « 75 ans et + » (38% ; n = 17) et le plus de personnes résidant dans des communes « Rural, défavorisé » et « Rural, intermédiaire » (13% et 13%, n = 6 et n = 6).

### 3.6.2 Analyse de la survie

La fonction de survie globale pour la modalité « limité » présente un effectif initial de 597 individus ( $T_0$  = date du diagnostic). Cinq ans après le diagnostic le nombre d'individus encore à risque (non perdu de vue et non décédé) est de 51. Le nombre d'évènement observé sur la période, ici le décès, est de 63, les autres individus étant perdus de vue. La durée de survie médiane est supérieure à la durée du suivi (5 ans). La probabilité de survie globale à cinq ans pour un cancer du sein « limité » en Corse selon la méthode de Kaplan-Meier est de 62,6% IC95[54,9 ; 71,3].

La fonction de survie globale pour la modalité « évolué » présente un effectif initial de 287 individus ( $T_0$  = date du diagnostic). Cinq ans après le diagnostic le nombre d'individus encore à risque est de 35. Le nombre d'évènement observé sur la période est de 72, les autres individus étant perdus de vue. La durée de survie médiane est de 1714 jours soit environ quatre ans et huit mois. La probabilité de survie globale à cinq ans pour un cancer du sein « évolué » en Corse selon la méthode de Kaplan-Meier est de 47,3% IC95[38,9 ; 57,6].

Les deux courbes de survie ne se croisent pas durant la période d'observation. La p-value du test de log-rank effectué est inférieur à la valeur seuil 0,05 (p-value <0,01). L'hypothèse  $H_0$  est donc rejetée : il existe une différence de survie en fonction du stade au diagnostic au risque alpha = 5%. La probabilité de survie à cinq ans est donc inférieure pour les individus dont le stade au diagnostic est « évolué » par rapport à ceux dont le stade au diagnostic est « limité ».

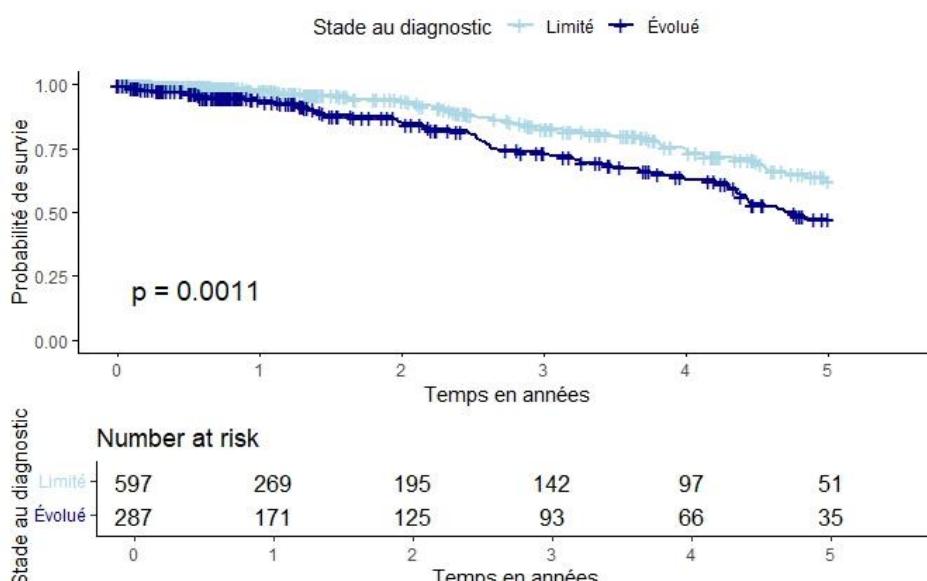


Figure 5 : Courbes de survie globale à cinq ans calculées en fonction du stade au diagnostic avec la méthode de Kaplan-Meier, valeur de la p-value du test du log-rank et tableau du nombre d'individus à risque aux différents temps.

N.B. : Sont disponibles en Annexe 22 les courbes, p-value et « risk set » des analyses de survie globale réalisées sur le modèle utilisé ci-dessus en fonction du stade au diagnostic en quatre catégories et des variables pour lesquelles une association brute a été mise en évidence avec le stade au diagnostic.

## 4 Discussion

### 4.1 Résultats clés

Cette étude est une première en termes de description du cancer du sein chez la femme en Corse à partir de données observées : aspect diagnostic, clinique et socio-démographique.

La proportion de cancers du sein à un stade « limité » sur l'île est plus importante qu'en France métropolitaine selon les dernières données publiées par SpF portant sur la période 2009-2012 [3]. Plus de la moitié des patientes atteintes d'un cancer du sein ont entre 50 et 75 ans et sont donc éligibles au programme de dépistage organisé mais seulement 29% des cancers de patientes de cette tranche d'âge sont détectés à la suite de la participation de la patiente à un dépistage organisé.

À l'aide de test du Chi<sup>2</sup> plusieurs associations brutes entre le stade au diagnostic en deux catégories et les variables disponibles ont pu être mises en évidence. Les circonstances du diagnostic influencent le stade au diagnostic. En effet la détection des cancers via le dépistage organisé permet de détecter plus souvent des cancers à un stade « limité » (23% vs 9% de stade évolué lors du dépistage organisé). À contrario, lorsque la détection se fait par le patient seul 19% des cancers ont un stade « limité » au diagnostic (vs 38% un stade « évolué »). Une association brute a également été mise en évidence entre la classe d'âge au diagnostic et le stade. Les personnes dont le stade est « limité » sont 60% à être dans la classe d'âge éligible au dépistage organisé vs 53% pour les stades « évolué ». Toutes les autres catégories d'âge, notamment les plus jeunes, sont surreprésentés dans les stades « évolué » vs stade « limité ». La dernière association brute mise en évidence concerne le département de résidence. Les personnes résidant en Haute-Corse sont plus nombreuses à avoir un stade au diagnostic « limité » que celles en Corse-du-Sud (56% vs 44%). Il est à noter qu'aucune association brute entre les variables sociodémographiques et le stade au diagnostic n'a pu être mise en évidence contrairement à d'autres études précitées [5,7].

Lors de l'analyse multivariée, que ce soit en analyse en cas complet ou sur données imputées, seules les circonstances de la découverte conservent une association au risque alpha = 5% avec le stade au diagnostic. La modalité de référence dans le modèle étant le dépistage organisé on constate dans les deux modèles une augmentation de la cote de découverte d'un cancer à un stade « évolué » lors d'une découverte de la tumeur par le médecin de 2,25 IC95[1,40 ; 3,73] lors de l'analyse en cas complet et 2,06 IC95[1,27 ; 3,34] lors de l'analyse avec données imputées. Cette augmentation de la cote de découverte d'une tumeur à un stade « évolué » est encore plus prononcée lorsque la patiente est à l'origine de la découverte : cote multipliée par 5,01 IC95[2,97 ; 8,69] lors de l'analyse en cas complet et 4,18 IC95[2,43 ; 7,18] lors de l'analyse avec données imputées.

Un modèle plus parcimonieux a été défini en se basant sur le critère AIC, à la fois lors de l'analyse en cas complet et avec les données imputées : pour les deux analyses seules les circonstances de la découverte ont été retenues. Pour les deux analyses, les RC ainsi que leurs intervalles de confiance à 95% étaient similaires à ceux du modèle intégrant l'ensemble des covariables.

En analysant les circonstances de la découverte en sous-groupes il a été détecté des différences au sein des groupes « Médecin » et « Patient ». La proportion d'individus avec un stade au diagnostic « évolué » est plus importante parmi les patientes dont les circonstances de la découverte est « Dépistage individuel » par rapport à celles ayant participé au programme de dépistage organisé (28% vs 16%), cette différence est d'autant accentuée quand on regarde les stades métastatiques (4,5% vs 0,7%). L'autopalpation représente 77% des circonstances détaillées de la découverte regroupée en « Patient ». Les femmes de moins de 40 ans sont surreprésentées dans cette catégorie par rapport à

l'ensemble des patientes (8,8% vs 3,9). Tout comme pour le dépistage organisé ou le dépistage individuel, la classe communale « Urbain, favorisé » est surreprésentée dans cette catégorie. Le stade au diagnostic est dans 49% des cas « évolué », ce qui en fait les circonstances de la découverte la deuxième plus défavorable juste après la « Rétractation mamelonnaire » avec 58% des cas avec un stade « évolué » au moment du diagnostic. La catégorie « Rétractation mamelonnaire » se caractérise avec la proportion de personnes de plus de 75 ans la plus élevée (38%) et la plus forte proportion d'individu résidant dans des communes appartenant à la classe communale « Rural, défavorisé » (13%) et « Rural, intermédiaire » (13%).

L'analyse de la survie a montré une différence statistiquement significative de la survie globale à cinq ans (T0 = date de diagnostic) des femmes atteintes d'un cancer du sein selon le stade au diagnostic. La probabilité de survie globale à cinq ans est plus faible pour les femmes dont le stade au diagnostic est « évolué » par rapport à celle qui ont un stade au diagnostic « limité ». La médiane de survie des femmes ayant un cancer à un stade au diagnostic « limité » est supérieure à cinq ans alors que celui des femmes ayant un cancer à un stade au diagnostic « évolué » est d'environ quatre ans et huit mois.

## 4.2 Limitations

L'incertitude sur les circonstances réelles de la découverte pose une limite importante de cette étude. Les circonstances de la découverte ne sont pas une variable obligatoire dans le cadre d'un registre des cancers et ce projet d'étude n'existe pas encore lors de la collecte des données 2017 à 2019. Elles n'ont donc pas été recherchées spécifiquement dans le cadre de la collecte des données en routine pour les patientes. Les circonstances de la découverte ont été déduites à partir des données disponibles par la relecture des dossiers. Pour certaines patientes par exemple, la notion de dépistage a été renseignée dans la case correspondante aux circonstances de la découverte dans le compte rendu de RCP sans précision s'il s'agissait d'un dépistage organisé ou d'un dépistage individuel, en cas de doute elles ont été classées en dépistage individuel sans preuve. Parfois, différentes circonstances de la découverte ont été renseignées lors de différentes RCP, le choix de celle conservée in-fine s'est fait selon les informations disponibles dans les autres données et selon la conviction de la personne chargée du codage. Dernière incertitude concernant les circonstances de la découverte est le caractère réellement connu de celle-ci de la part de l'équipe médicale et donc renseignée par celle-ci. Il est imaginable que certaines patientes aient palpé une tumeur dans leur sein, soient allé voir leur médecin traitant qui a posé le diagnostic, sans pour autant reporter le fait que ce soit la patiente qui ait perçu les signes d'alertes menant à l'examen clinique ou au dépistage individuel.

La méconnaissance de facteurs de risques individuels de patientes qu'ils soient liés à des facteurs médicaux (génétiques, troubles psychiatriques, ...), comportementaux (retard de consultation après détection d'un signe d'alerte par la patiente, ...) ou d'habitude de vie (obésité, consommation de substances stupéfiantes, ...) sont également un biais de cette étude. Il est tout à fait imaginable que ces facteurs de risques amènent à un retard de diagnostic au-delà des circonstances de la découverte et jouent un rôle important dans la découverte à un stade « évolué » du cancer du sein.

L'utilisation de données agrégées au niveau communal peut être considérée comme une limite de l'étude même si de nombreuses études utilisent ce procédé [5,7,30,31] et que des études ont démontré l'utilité de ce genre de procédé, l'individu étant influencé par son environnement [32,33]. Le niveau géographique aurait éventuellement pu être réduit en utilisant les Ilots Regroupés pour l'Information Statistique (IRIS) au lieu du code commune de résidence des patientes. Les adresses de résidence étant cependant régulièrement

renseignées de façon partielle, un balayage individuel aurait été nécessaire afin de s'assurer de l'exactitude des codes IRIS renseignés.

S'ajoute au choix du niveau de découpe géographique la difficulté des « communes éclatées » [21]. Au nombre de 42 en 2017, ce sont des communes pour lesquelles la majorité des habitants résident hors du chef-lieu. Pour ces communes les temps de trajet au centre de mammographie le plus proche n'est pas forcément pertinent car celui-ci est calculé de mairie à mairie. Le problème a cependant été atténué quand le code postal du lieu de résidence est différent du code postal du chef-lieu et qu'une mairie annexe existe au code postal de résidence.

D'une façon générale l'ensemble des regroupements peut être remis en question. Il s'agissait dans l'étude de trouver un compromis entre la précision des données dont nous disposions et la création de regroupements permettant aux acteurs en santé insulaire de se saisir des résultats afin d'améliorer la prise en charge du cancer du sein des femmes résidant sur le territoire.

La faiblesse des effectifs, 814 individus pour l'étude en cas complet et 884 individus pour l'étude contenant des données imputées, n'a peut-être pas permis de démontrer un effet statistiquement significatif de certains facteurs de risque d'un stade « évolué » au diagnostic d'un cancer du sein. Il est possible qu'une cohorte plus importante, une durée d'inclusion plus longue ou un seuil de significativité supérieur nous aurait permis de tirer des conclusions qui ont été occultées lors de cette étude.

L'analyse de la survie repose uniquement sur les données de mortalité connues par le RG2C via ses différentes sources d'informations collectées en routine et d'un croisement réalisé à partir du fichier du RNIPP. Afin d'améliorer l'exhaustivité des décès ou de repousser les dates de dernière nouvelle, une consultation systématique des mairies de naissance des individus, des mairies de résidence, des caisses primaires d'assurance maladie et du dossier médical pourrait diminuer le nombre d'individus perdus de vue.

Les causes de décès ne sont pas connues par le RG2C. Cette absence d'information introduit un biais d'information, les décès non dus au cancer étant également comptabilisé. L'analyse ne permet pas de déduire une surmortalité liée au cancer (vs population générale), une analyse de la survie nette étant nécessaire afin de comparer la population de patiente porteuse d'une pathologie tumorale à une cohorte de patiente aux mêmes caractéristiques.

### 4.3 Interprétation

La proportion de patientes avec un stade « limité » est plus importante en Corse qu'en France métropolitaine selon les dernières données disponibles. Il est cependant nécessaire de relativiser ce constat du fait d'une comparaison doublement biaisée. L'étude de SpF porte sur une période différente s'étalant de 2009 à 2012 alors que cette étude inclus des cas incident entre 2017 et 2019. Il est possible que les techniques de dépistage et l'adhésion du public aux campagnes de dépistage (organisé ou non) aient évolué dans cet intervalle de temps, notamment à la suite du plan cancer 2009 – 2013 dont l'axe 3 est dédié à la prévention et au dépistage [34]. Le second point de vigilance concerne la différence de dénombrement des cancers à un stade avancé. Le rapport de SpF comptabilise les cancers non réséqués en stade « Avancé », pour des raisons pronostiques. L'approche de cette étude est différente. Nous avons considéré les tailles tumorales au diagnostic et leur propagation sans prendre en considération des choix thérapeutiques qui peuvent être individuels et non en lien direct avec la tumeur.

Cette étude a toutefois mis en évidence l'intérêt du dépistage organisé pour la détection précoce des cancers du sein. Dans l'analyse multivariée, les circonstances de la découverte est l'unique variable ayant un effet statistiquement significatif sur le stade au diagnostic en

deux modalités. L'écart observé entre les différentes modalités reste cependant à être confirmé par des études similaires dans d'autres aires géographiques ou par une amélioration du système de classement des circonstances de la découverte.

La forte proportion des stades « limité » parmi les découvertes de cancers à la suite du dépistage organisé doit faire réfléchir à l'opportunité d'élargir la classe d'âge de dépistage des cancers, notamment aux individus plus jeunes, comme c'est déjà le cas dans certains pays [35]. Le bénéfice sanitaire et économique d'une détection précoce serait élevé en évitant à la fois des opérations lourdes et mutilantes pour ces femmes et en diminuant la mortalité globale. Cette question est d'autant plus importante que la part d'individus dans la classe d'âge « [40 ; 50[ ans » représente plus de 10% des cancers incidents, dont 40% de stade évolué (vs 32% en population générale). L'élargissement du dépistage organisé augmente cependant le risque de surdiagnostic même si cette théorie rencontre de nombreuses controverses [36].

En examinant les circonstances de la découverte détaillées, deux éléments sont particulièrement marquants. Le premier est la proportion de stade au diagnostic « évolué » plus important parmi les patientes dont la découverte s'est faite via le dépistage individuel que via le dépistage organisé (28% vs 16%). Cet écart plaide en faveur du dépistage organisé qui permet, selon les sources, de détecter entre 6% et 9% des cancers lors de la seconde lecture non détectés à la première [37,38]. Ce constat est cependant à relativiser au vu de la faiblesse de la certitude de la circonstance de la découverte détaillée. D'une part les individus pour lesquels seule la notion de dépistage était connue sans savoir s'il s'agissait d'un dépistage individuel ou organisé ont été classés en dépistage individuel entraînant un potentiel biais de classement. D'autre part, certains individus ont potentiellement eu des symptômes ou des doutes les amenant à consulter leur médecin traitant, qui a prescrit une mammographie de dépistage sans que l'information sur les motifs de la prescription soient connue par l'équipe du RG2C. Le second élément marquant concerne la proportion de stade « évolué » parmi les personnes dont le diagnostic a été posé à la suite d'une autopalpation. L'autopalpation est souvent évoquée comme moyen de détection précoce du cancer du sein, et semble connue des femmes françaises comme tel [39,40]. La haute autorité de santé (HAS) ne recommande cependant pas l'autopalpation [38] car aucun bénéfice de l'autopalpation n'a pu être démontré en population générale [41]. Cela pourrait s'expliquer par un manque de régularité et de connaissance des techniques de réalisation [42] aboutissant ainsi à une part de stade « évolué » au diagnostic plus importante qu'attendu initialement.

Contrairement à d'autres études précitées [5,7,11] il n'a pas été mis en évidence de lien statistiquement significatif entre le stade au diagnostic évolué et les facteurs socio-économiques dans notre étude. L'hypothèse optimiste est une répartition plus uniforme de la pauvreté sur le territoire insulaire, évitant ainsi une « ghettoïsation » de certaines communes comme cela peut-être le cas en Ile-de-France ou ailleurs sur le territoire. L'absence de données individuelles mais uniquement à l'échelle communale gommerait ainsi les distinctions sociodémographiques entre les différentes patientes de l'étude. Une seconde hypothèse, possiblement plus réaliste, est la taille insuffisante de l'échantillon pour mettre en évidence une association significative au risque alpha de 5% entre le stade au diagnostic et les indicateurs sociodémographiques retenus. En considérant les intervalles de confiance des modalités « Rural, défavorisé » et « Rural, intermédiaire » de la variable classe communale, nous constatons un désaxement de ses modalités dans le sens d'une augmentation du risque par rapport à la modalité de référence qui est « Urbain, favorisé ». Ainsi il est possible qu'en augmentant la taille de l'échantillon une différence statistiquement significative puisse être mesurée sur le territoire insulaire tout comme cela a été le cas dans d'autres études.

Une association brute, statistiquement significative, a été mise en évidence entre le département de résidence et le stade au diagnostic, tant lors de l'analyse en cas complet que sur données imputées. Cette association brute, dans le sens d'une plus grande proportion de cancer à un stade « évolué » lors du diagnostic en Corse-du-Sud qu'en Haute-Corse, nécessite de s'interroger sur les différences départementales conduisant à cet écart. Il serait intéressant de conduire d'autres études dans le contexte afin de confirmer la différence et d'en déterminer les causes.

#### 4.4 Validité externe des résultats

Cette étude est la première sur la thématique menée sur le territoire insulaire à partir de données issues du RG2C et se basant ainsi sur des données mesurées et non estimées. L'exhaustivité de données présentées renforce la validité interne des résultats.

Cependant les spécificités insulaires (forte défavorisation sociale, statut d'île-montagne, faible densité de population, ...) peuvent limiter leur généralisation à d'autres régions. Il serait désormais pertinent de renouveler cette étude sur des territoires similaires en France à partir de données des registres des cancers existant dans d'autres départements ou dans d'autres territoires méditerranéens afin de pouvoir confronter les résultats.

La réutilisation de données préexistantes pour notre étude n'a pas permis d'inclure certaines variables cliniques (fiabilité incontestable des circonstances de la découverte), comportementale (statut tabagique, alcoolisme, obésité, ...) ou sociale (défavorisation sociale individuelle) susceptibles d'influencer le stade au diagnostic. Une étude complémentaire, avec recours systématique au dossier patient, pourrait enrichir la compréhension des facteurs influençant la probabilité d'un stade au diagnostic « évolué ». L'enrichissement des données avec des données issues des centres régionaux de coordination du dépistage des cancers pourrait également améliorer la fiabilité des résultats et ainsi améliorer la validité externe des résultats.

Cette étude a été réalisée sur des données de la période pré-COVID afin de ne pas surévaluer la part de stade au diagnostic « évolué » due à d'éventuels retard au diagnostic lié à la pandémie. Cette période et notamment les confinements ont modifié les comportements et les perceptions de la population en termes de santé. Il serait intéressant de renouveler l'étude sur une période post-COVID afin de mesurer d'éventuelles différences comportementales de recours aux soins entraînant des différences de délais de diagnostic et modifiant le stade de la maladie lors de son diagnostic.

## 5 Conclusion

Cette étude novatrice du fait de sa zone d'étude et du croisement de données médicales, individuelles et écologiques (agrégées à la commune de résidence) permet d'ouvrir la voie à des études scientifiques dans le domaine de la santé publique en région Corse à partir des données du RG2C.

Cette étude dresse un panorama des cancers du sein incident entre 2017 et 2019 sur le territoire insulaire, permettant à la fois de mieux connaître leur répartition selon des critères :

- Médicaux : la morphologie tumorale, le statut hormonal et le stade au diagnostic ;
- Individuels : la répartition des patientes selon leur âge, les circonstances ayant mené au diagnostic de la tumeur et aspects comportementaux ;
- Environnementaux : critères médico-sociaux des communes de résidence des patientes et temps de trajet au centre de dépistage le plus proche.

L'étude a permis de mettre en évidence l'importance des circonstances de la découverte de la maladie dans la probabilité de découverte à un stade « évolué ». Elle renouvelle ainsi la démonstration de l'importance du dépistage organisé dans la détection des cancers du sein à un stade « limité ». L'augmentation du risque de découverte d'un cancer à un stade « évolué » a été mesurée lors d'une découverte par le médecin ou la patiente en comparaison à une découverte lors du dépistage organisé. Une tendance, documentée dans d'autres études, a également pu être observé en faveur du dépistage organisé comparativement au dépistage individuel ou à l'autopalpation.

Contrairement à d'autres études l'effet des indicateurs environnementaux n'a pas pu être mis en évidence même si une tendance d'augmentation du risque a pu être observée chez des patientes résidant dans des communes classées comme « Rural, défavorisé » et « Rural, intermédiaire » par rapport aux patientes résidant dans des communes « Urbain, favorisé ».

Afin de servir au mieux à la population insulaire cette étude doit désormais être portée à la connaissance des acteurs régionaux en santé publique afin qu'ils puissent se saisir des résultats présentés. Une diffusion via le site internet de l'observatoire régional de la santé de Corse ([www.ors.corsica](http://www.ors.corsica)) est prévue. Une présentation de l'étude et des résultats pourra être envisagée auprès des acteurs régionaux qui en feront la demande. L'opportunité de rédiger, à partir des résultats de l'étude, un article ayant vocation à être publié dans une revue scientifique à comité de lecture sera prochainement étudié par le médecin responsable de la recherche du RG2C.

## Annexes

### Annexe 1. Répartition des tumeurs selon le FDep et l'EDI

**Répartition des tumeurs selon le FDep  
de la commune de résidence du patient**

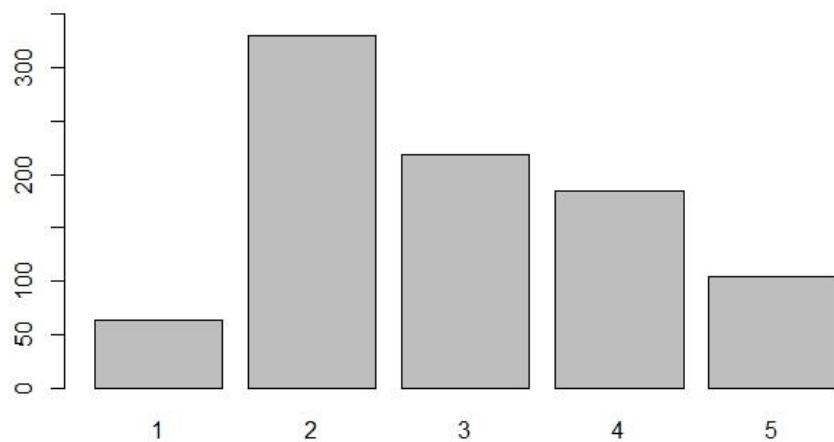


Figure 6 : Répartition du nombre de tumeurs en fonction du FDep de la commune de résidence du porteur de la pathologie

**Répartition des tumeurs selon l'EDI  
de la commune de résidence du patient**

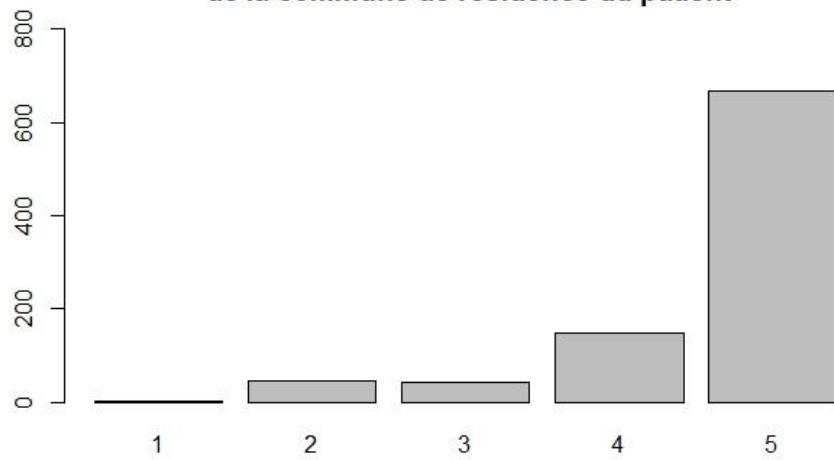


Figure 7 : Répartition du nombre de tumeurs en fonction de l'EDI de la commune de résidence du porteur de la pathologie

## Annexe 2. Informations de la session R

```
R version 4.4.2 (2024-10-31 ucrt)
Platform: x86_64-w64-mingw32/x64
Running under: Windows 11 x64 (build 26100)

Matrix products: default

locale:
[1] LC_COLLATE=French_France.utf8  LC_CTYPE=French_France.utf8    LC_MONETARY=French_France.utf8
[4] LC_NUMERIC=C                  LC_TIME=French_France.utf8

time zone: Europe/Paris
tzcode source: internal

attached base packages:
[1] grid      stats     graphics  grDevices utils      datasets  methods    base

other attached packages:
[1] survminer_0.5.0   ggpubr_0.6.0     survival_3.7-0    stringi_1.8.4   miceadds_3.17-44  broom_1.0.8
[7] reshape2_1.4.4    vcd_1.4-13     dendextend_1.19.0 purrr_1.0.4     stringr_1.5.1     sf_1.0-20
[13] factoextra_1.0.7 FactoMiner_2.11  cluster_2.1.6    ggplot2_3.5.2   mice_3.17.0      writexl_1.5.4
[19] forcats_1.0.0    flextable_0.9.7  gtsummary_2.2.0  tidyverse_1.3.1  dplyr_1.1.4

loaded via a namespace (and not attached):
[1] Rdpack_2.6.4          DBI_1.2.3          gridExtra_2.3      rlang_1.1.5
[5] magrittr_2.0.3         e1071_1.7-16       compiler_4.4.2     systemfonts_1.2.2
[9] vctrs_0.6.5           pkgconfig_2.0.3    shape_1.4.6.1     fastmap_1.2.0
[13] backports_1.5.0       KMSurv_0.1-6       rmarkdown_2.29     nloptr_2.2.1
[17] ragg_1.4.0            xfun_0.51          glmnet_4.1-8      jomo_2.7-6
[21] flashClust_1.01-2    uuid_1.2-1         pan_1.9           R6_2.6.1
[25] car_3.1-3             boot_1.3-31       rpart_4.1.23      lmtest_0.9-40
[29] estimability_1.5.1   Rcpp_1.0.14        iterators_1.0.14  knitr_1.49
[33] zoo_1.8-14            Matrix_1.7-1       splines_4.4.2     nnet_7.3-19
[37] tidyselect_1.2.1      abind_1.4-8        viridis_0.6.5     codetools_0.2-20
[41] plyr_1.8.9             lattice_0.22-6     tibble_3.2.1      withr_3.0.2
[45] askpass_1.2.1          evaluate_1.0.3    units_0.8-7      proxy_0.4-27
[49] zip_2.3.2              xml2_1.3.8        survMisc_0.5.6    pillar_1.10.2
[53] carData_3.0-5          KernSmooth_2.23-24 DT_0.33          foreach_1.5.2
[57] reformulas_0.4.0       generics_0.1.3    munsell_0.5.1     scales_1.3.0
[61] minqa_1.2.8            xtable_1.8-4       leaps_3.2         class_7.3-22
[65] glue_1.8.0              gdtools_0.4.2     emmeans_1.11.0    scatterplot3d_0.3-44
[69] tools_4.4.2             data.table_1.17.0  lme4_1.1-37       ggsignif_0.6.4
[73] mvtnorm_1.3-3          mitools_2.4       rbibutils_2.3     colorspace_2.1-1
[77] nlme_3.1-166            Formula_1.2-5     cli_3.6.4         km.ci_0.5-6
[81] textshaping_1.0.0       officer_0.6.8     fontBitstreamVera_0.1.1 viridisLite_0.4.2
[85] gtable_0.3.6            rstatix_0.7.2     digest_0.6.37     fontquiver_0.2.1
[89] classInt_0.4-11          ggrepel_0.9.6     htmlwidgets_1.6.4  htmltools_0.5.8.1
[93] lifecycle_1.0.4          multcompview_0.1-10 mitml_0.4-5       fontLiberation_0.1.0
[97] openssl_2.3.2           MASS_7.3-61
```

Figure 8 : Informations de la session R utilisée pour l'exploitation statistique des données obtenus via la commande `sessionInfo()`.

### Annexe 3. Choix du nombre d'imputations

m	DO	Medecin	Patient
1	0,1900452	0,5565611	0,2533937
2	0,1798643	0,5656109	0,2545249
3	0,1840121	0,5603318	0,2556561
5	0,1873303	0,5574661	0,2552036
10	0,1842760	0,5609729	0,2547511
15	0,1840875	0,5609351	0,2549774
20	0,1852941	0,5585973	0,2561086
50	0,1840724	0,5605204	0,2554072
100	0,1848077	0,5605995	0,2545928

Tableau 3 : Evolution des proportions des modalités de la variable circonstance de la découverte imputées en fonction du nombre d'imputations (m) réalisées

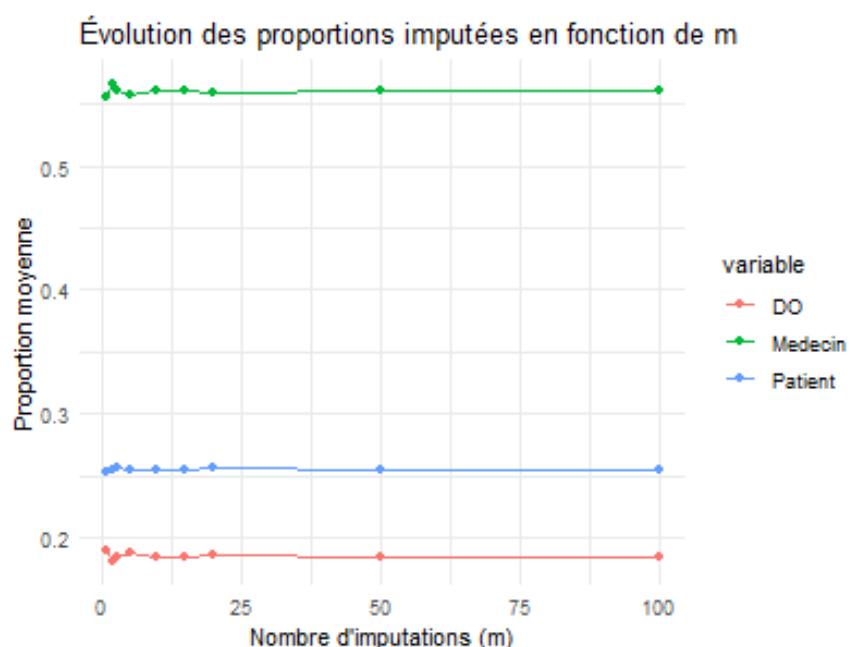


Figure 9 : Représentation graphique de l'évolution des proportions des modalités de la variable circonstance de la découverte imputées en fonction du nombre d'imputations (m) réalisées

#### Annexe 4. Ecriture mathématique du modèle logistique présenté

Le modèle logistique proposé est de la forme suivante :

$$\text{logit}(P(Y_i = 1|X_i)) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_n X_{in} + \varepsilon_i$$

*Équation 1 : Formule générale du modèle logistique proposé afin de déterminer les coefficients des variables prédictives d'un stade au diagnostic évolué*

Avec :

- $Y_i$  : Le stade au diagnostic de l'individu i
- $\beta_0$  : Le terme constant
- $\beta_n$  : Le coefficient de la variable n
- $X_{in}$  : La valeur de la variable n pour l'individu i
- $\varepsilon_i$  : L'erreur aléatoire associée à l'individu i

## Annexe 5. Diagramme de flux

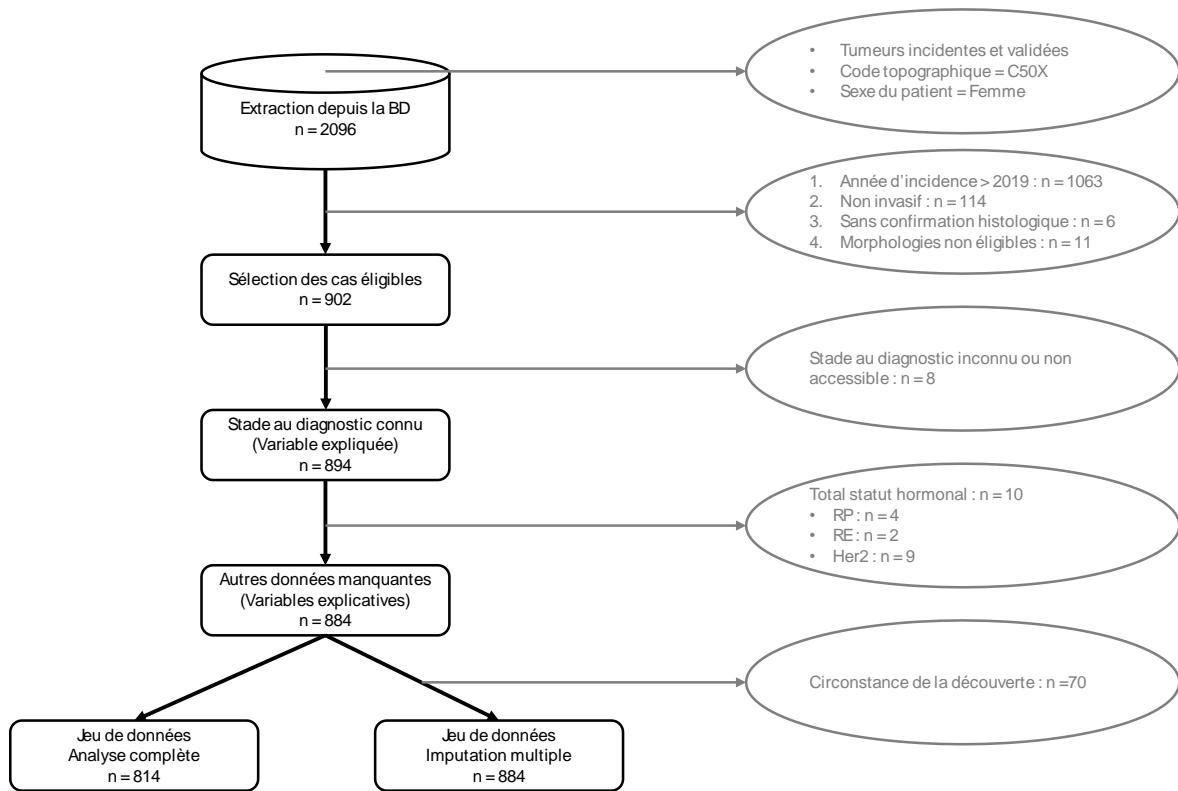


Figure 10 : Diagramme de flux des données utilisées pour l'étude

## Annexe 6. Description du jeu de données, avant imputation

Caractéristiques	N = 884 <sup>1</sup>
<b>Age</b>	26 - 54 - 65 - 65 - 75 - 98
<b>Stade ENCR</b>	
Local limite	587 / 884 (66%)
Local étendu	10 / 884 (1.1%)
Régional	234 / 884 (26%)
Avancé	53 / 884 (6.0%)
<b>Circonstance de découverte – 3 catégories</b>	
DO	150 / 814 (18%)
Médecin	458 / 814 (56%)
Patient	206 / 814 (25%)
<b>Circonstance de découverte – 7 catégories</b>	
Autopalpation	148 / 814 (19%)
GG axillaire	4 / 814 (0.5%)
Rétraction, ...	45 / 814 (6%)
Examen clinique	117 / 814 (14%)
Mammographie	289 / 814 (36%)
Bilan	54 / 814 (6.6%)
Dépistage organisé	150 / 814 (19%)
<b>FDep (quintiles)</b>	
1 (Très favorisé)	63 / 884 (7.1%)
2	326 / 884 (37%)
3	211 / 884 (24%)
4	183 / 884 (21%)
5 (Très défavorisé)	101 / 884 (11%)
<b>Département de résidence</b>	
Corse-du-Sud	421 / 884 (48%)
Haute-Corse	463 / 884 (52%)

<sup>1</sup>Min - Q1 - Mean - Median - Q3 - Max; n / N (%)

<b>Caractéristiques</b>	<b>N = 884<sup>1</sup></b>
<b>Temps (continu)</b>	0 - 0 - 14 - 9 - 26 - 86
<b>Temps</b>	
Commune avec Mammo	394 / 884 (45%)
Moins de 15min	145 / 884 (16%)
Entre 15 et 30min	180 / 884 (20%)
Plus de 30min	165 / 884 (19%)
<b>Baccalauréat</b>	
Non	290 / 884 (33%)
Oui	594 / 884 (67%)
<b>APL</b>	
Inférieur à la moyenne	231 / 884 (26%)
Supérieur à la moyenne	653 / 884 (74%)
<b>Statut hormonal</b>	
Basal like	73 / 884 (8.3%)
Her2 surexprimé	54 / 884 (6.1%)
Luminal	757 / 884 (86%)

<sup>1</sup>Min - Q1 - Mean - Median - Q3 - Max; n / N (%)

Tableau 4 : Caractéristiques cliniques, sociodémographiques et diagnostiques des patientes atteintes de cancer du sein, jeu de données avant imputation

**Annexe 7. Répartition des patientes selon leur âge et la distance isochrone entre leur commune de résidence et le centre de mammographie le plus proche, avant imputation**

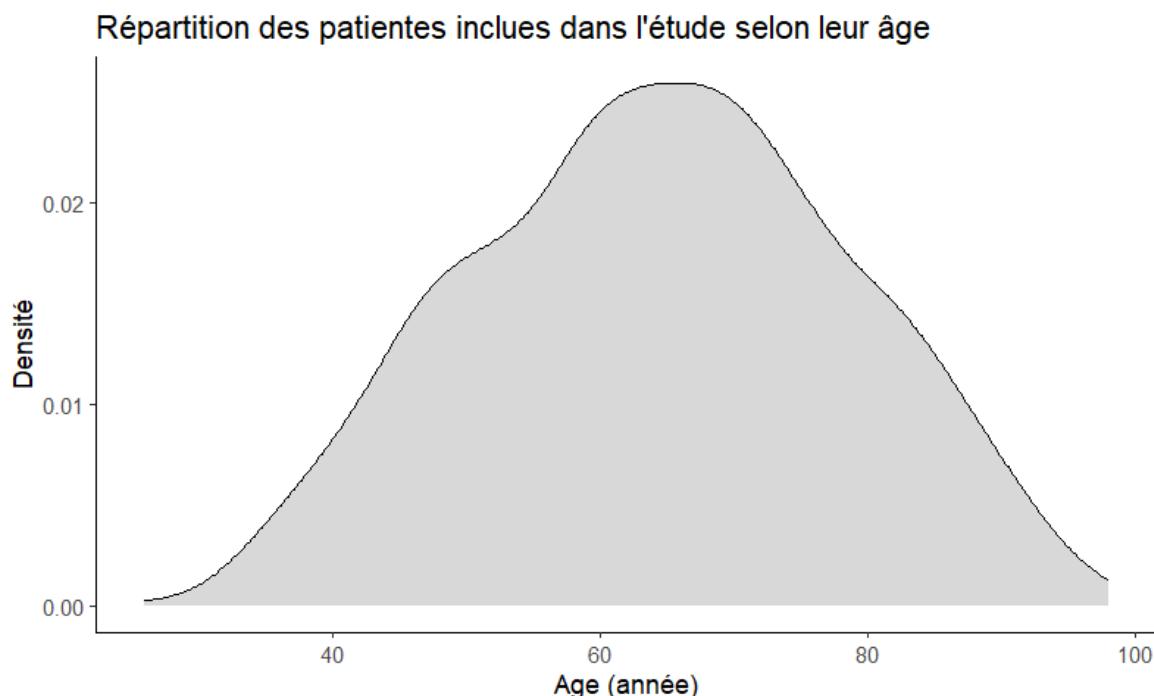


Figure 11 : Densité de répartition des patientes porteuses d'un cancer du sein selon leur âge, avant imputation

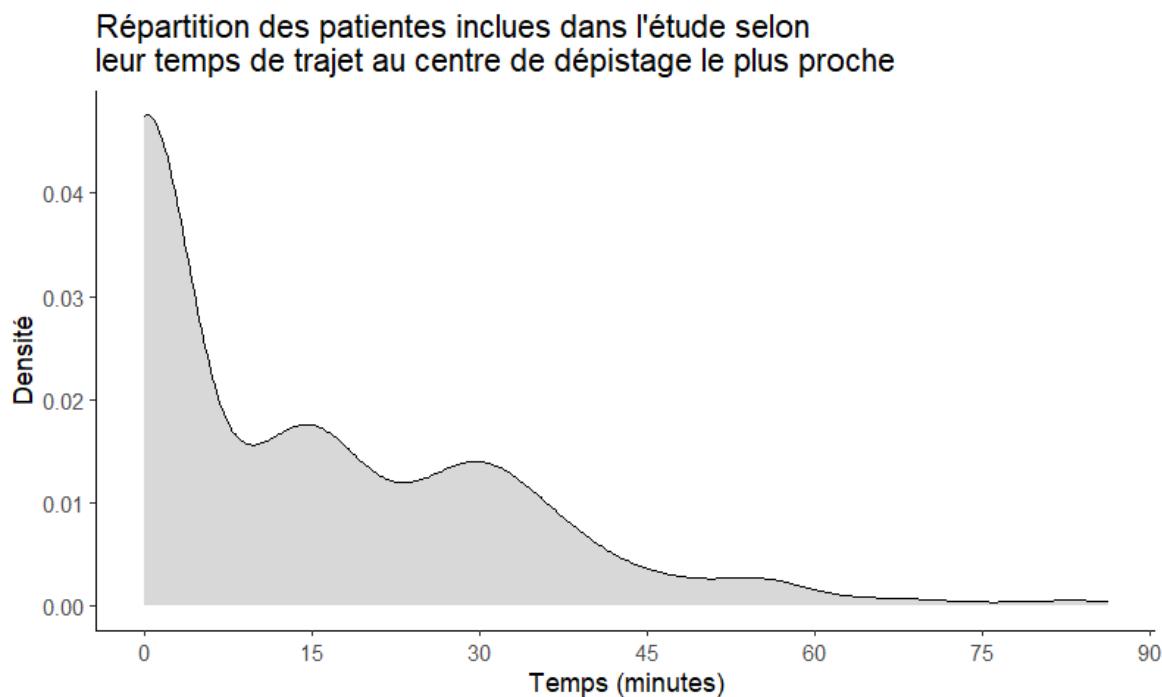


Figure 12 : Densité de répartition des patientes porteuses d'un cancer du sein selon la distance isochrone entre leur commune de résidence et le centre de mammographie le plus proche, avant imputation

## Annexe 8. Description du jeu de données, « cas complet »

Caractéristiques	N = 814 <sup>1</sup>
<b>Age</b>	26 - 54 - 65 - 65 - 75 - 98
<b>Stade ENCR</b>	
Local limite	538 / 814 (66%)
Local étendu	10 / 814 (1.2%)
Régional	215 / 814 (26%)
Avance	51 / 814 (6.3%)
<b>Circonstance de découverte – 3 catégories</b>	
DO	150 / 814 (18%)
Médecin	458 / 814 (56%)
Patient	206 / 814 (25%)
<b>Circonstance de découverte – 7 catégories</b>	
Autopalpation	148 / 814 (19%)
GG axillaire	4 / 814 (0.5%)
Rétraction, ...	45 / 814 (6%)
Examen clinique	117 / 814 (14%)
Mammographie	289 / 814 (36%)
Bilan	54 / 814 (6.6%)
Dépistage organisé	150 / 814 (19%)
<b>FDep (quintiles)</b>	
1 (Très favorisé)	56 / 814 (6.9%)
2	306 / 814 (38%)
3	187 / 814 (23%)
4	168 / 814 (21%)
5 (Très défavorisé)	97 / 814 (12%)
<b>Département de résidence</b>	
Corse-du-Sud	395 / 814 (49%)
Haute-Corse	419 / 814 (51%)

<sup>1</sup>Min - Q1 - Mean - Median - Q3 - Max; n / N (%)

<b>Caractéristiques</b>	<b>N = 814<sup>1</sup></b>
<b>Temps (continu)</b>	0 - 0 - 14 - 9 - 25 - 86
<b>Temps</b>	
Commune avec Mammo	367 / 814 (45%)
Moins de 15min	131 / 814 (16%)
Entre 15 et 30min	164 / 814 (20%)
Plus de 30min	152 / 814 (19%)
<b>Baccalauréat</b>	
Non	265 / 814 (33%)
Oui	549 / 814 (67%)
<b>APL</b>	
Inférieur à la moyenne	211 / 814 (26%)
Supérieur à la moyenne	603 / 814 (74%)
<b>Statut hormonal</b>	
Basal like	71 / 814 (8.7%)
Her2 surexprimé	48 / 814 (5.9%)
Luminal	695 / 814 (85%)

<sup>1</sup>Min - Q1 - Mean - Median - Q3 - Max; n / N (%)

**Annexe 9. Répartition des patientes selon leur âge et la distance isochrone entre leur commune de résidence et le centre de mammographie le plus proche, « cas complet »**

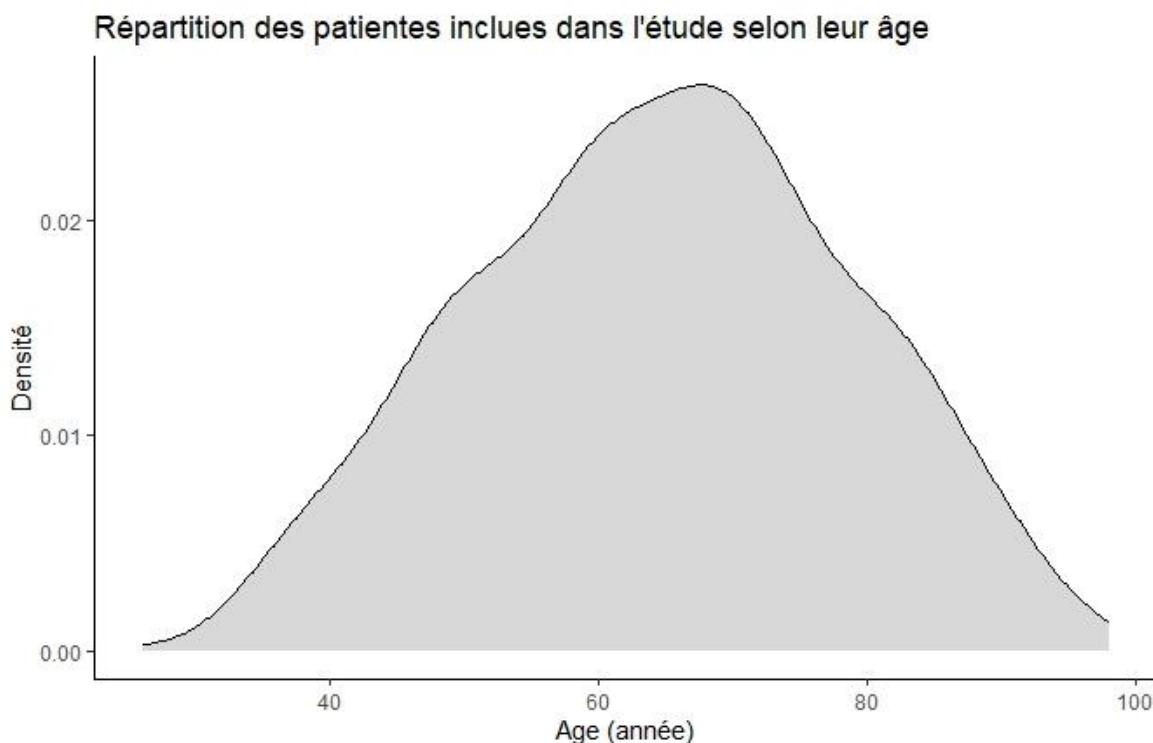


Figure 13 : Densité de répartition des patientes porteuses d'un cancer du sein selon leur âge, cas complet

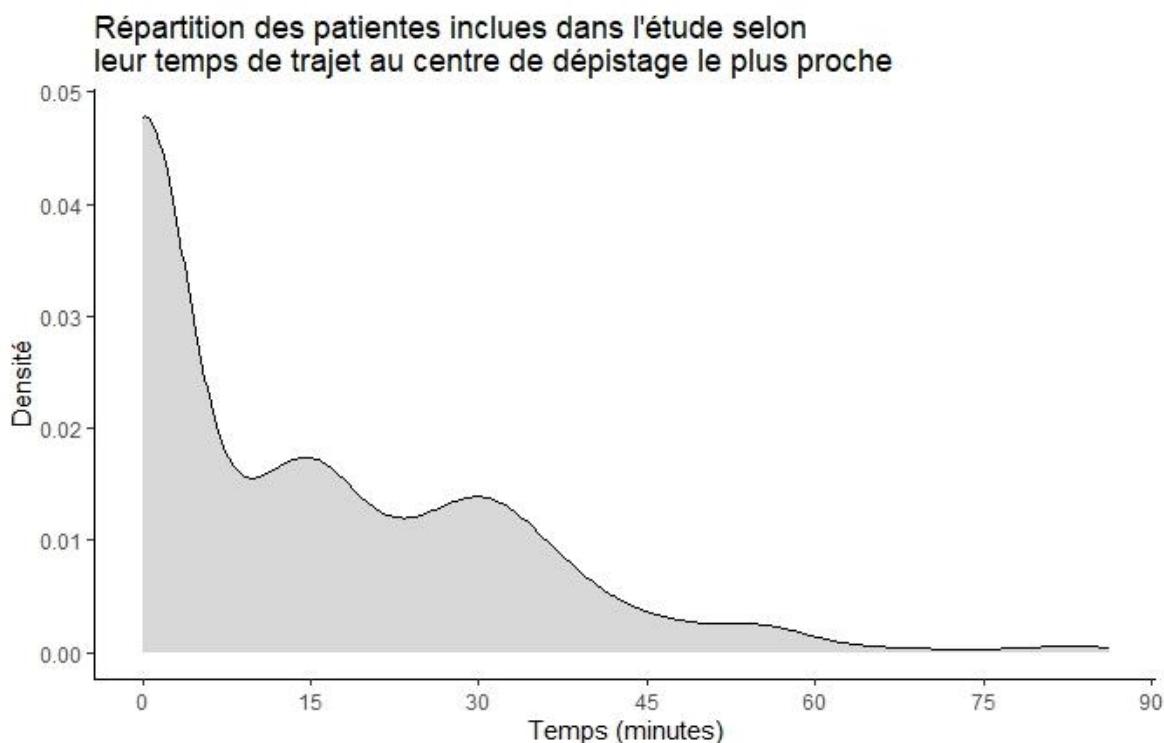


Figure 14 : Densité de répartition des patientes porteuses d'un cancer du sein selon la distance isochrone entre leur commune de résidence et le centre de mammographie le plus proche, avant imputation, cas complet

**Annexe 10.**
**Stade au diagnostic, jeu de données avant imputation**

<b>Caractéristiques</b>	<b>Local limite N = 587</b>	<b>Local étendu N = 10</b>	<b>Régional N = 234</b>	<b>Avance N = 53</b>
<b>Morphologie</b>				
Carcinome canalaire	442 / 587 (75%)	7 / 10 (70%)	176 / 234 (75%)	38 / 53 (72%)
Carcinome lobulaire	89 / 587 (15%)	3 / 10 (30%)	39 / 234 (17%)	7 / 53 (13%)
Carcinome SAI et autres	56 / 587 (9.5%)	0 / 10 (0%)	19 / 234 (8.1%)	8 / 53 (15%)
<b>Classe d'âge</b>				
[15-40[	17 / 587 (2.9%)	0 / 10 (0%)	15 / 234 (6.4%)	1 / 53 (1.9%)
[40-50[	69 / 587 (12%)	1 / 10 (10%)	39 / 234 (17%)	7 / 53 (13%)
[50-75[	360 / 587 (61%)	1 / 10 (10%)	127 / 234 (54%)	24 / 53 (45%)
75 et +	141 / 587 (24%)	8 / 10 (80%)	53 / 234 (23%)	21 / 53 (40%)
<b>Année de diagnostic</b>				
2017	202 / 587 (34%)	7 / 10 (70%)	69 / 234 (29%)	17 / 53 (32%)
2018	194 / 587 (33%)	2 / 10 (20%)	73 / 234 (31%)	20 / 53 (38%)
2019	191 / 587 (33%)	1 / 10 (10%)	92 / 234 (39%)	16 / 53 (30%)
<b>Circonstance de découverte – 3 catégories</b>				
DO	126 / 538 (23%)	0 / 10 (0%)	23 / 215 (11%)	1 / 51 (2.0%)
Médecin	312 / 538 (58%)	6 / 10 (60%)	106 / 215 (49%)	34 / 51 (67%)
Patient	100 / 538 (19%)	4 / 10 (40%)	86 / 215 (40%)	16 / 51 (31%)
<b>Statut hormonal</b>				
Basal like	48 / 587 (8.2%)	0 / 10 (0%)	20 / 234 (8.5%)	5 / 53 (9.4%)
Her2 surexprimé	30 / 587 (5.1%)	2 / 10 (20%)	19 / 234 (8.1%)	3 / 53 (5.7%)
Luminal	509 / 587 (87%)	8 / 10 (80%)	195 / 234 (83%)	45 / 53 (85%)
<b>FDep (quintiles)</b>				
1 (Très favorisé)	41 / 587 (7.0%)	0 / 10 (0%)	18 / 234 (7.7%)	4 / 53 (7.5%)
2	223 / 587 (38%)	5 / 10 (50%)	84 / 234 (36%)	14 / 53 (26%)
3	140 / 587 (24%)	0 / 10 (0%)	51 / 234 (22%)	20 / 53 (38%)
4	117 / 587 (20%)	3 / 10 (30%)	50 / 234 (21%)	13 / 53 (25%)
5 (Très défavorisé)	66 / 587 (11%)	2 / 10 (20%)	31 / 234 (13%)	2 / 53 (3.8%)

Caractéristiques	Local limite N = 587	Local étendu N = 10	Régional N = 234	Avance N = 53
<b>APL</b>				
Inférieur à la moyenne	158 / 587 (27%)	4 / 10 (40%)	54 / 234 (23%)	15 / 53 (28%)
Supérieur à la moyenne	429 / 587 (73%)	6 / 10 (60%)	180 / 234 (77%)	38 / 53 (72%)
<b>Baccalauréat</b>				
Non	190 / 587 (32%)	5 / 10 (50%)	77 / 234 (33%)	18 / 53 (34%)
Oui	397 / 587 (68%)	5 / 10 (50%)	157 / 234 (67%)	35 / 53 (66%)
<b>Temps</b>				
Commune avec Mammo	254 / 587 (43%)	4 / 10 (40%)	112 / 234 (48%)	24 / 53 (45%)
Moins de 15min	101 / 587 (17%)	1 / 10 (10%)	38 / 234 (16%)	5 / 53 (9.4%)
Entre 15 et 30min	119 / 587 (20%)	1 / 10 (10%)	46 / 234 (20%)	14 / 53 (26%)
Plus de 30min	113 / 587 (19%)	4 / 10 (40%)	38 / 234 (16%)	10 / 53 (19%)

Tableau 5 : Caractéristiques cliniques, sociodémographiques et diagnostiques des patientes atteintes de cancer du sein selon le stade d'extension au diagnostic, jeu de données avant imputation

**Annexe 11.**
**Stade au diagnostic, jeu de données « cas complet »**

Caractéristiques	Local limite N = 538 <sup>1</sup>	Local étendu N = 10 <sup>1</sup>	Régional N = 215 <sup>1</sup>	Avance N = 51 <sup>1</sup>
<b>Morphologie</b>				
Carcinome canalaire	398 / 538 (74%)	7 / 10 (70%)	165 / 215 (77%)	36 / 51 (71%)
Carcinome lobulaire	88 / 538 (16%)	3 / 10 (30%)	32 / 215 (15%)	7 / 51 (14%)
Carcinome SAI et autres	52 / 538 (9.7%)	0 / 10 (0%)	18 / 215 (8.4%)	8 / 51 (16%)
<b>Classe d'âge</b>				
[15-40[	17 / 538 (3.2%)	0 / 10 (0%)	14 / 215 (6.5%)	1 / 51 (2.0%)
[40-50[	57 / 538 (11%)	1 / 10 (10%)	37 / 215 (17%)	7 / 51 (14%)
[50-75[	333 / 538 (62%)	1 / 10 (10%)	116 / 215 (54%)	24 / 51 (47%)
75 et +	131 / 538 (24%)	8 / 10 (80%)	48 / 215 (22%)	19 / 51 (37%)
<b>Année de diagnostic</b>				
2017	199 / 538 (37%)	7 / 10 (70%)	68 / 215 (32%)	16 / 51 (31%)
2018	172 / 538 (32%)	2 / 10 (20%)	63 / 215 (29%)	19 / 51 (37%)
2019	167 / 538 (31%)	1 / 10 (10%)	84 / 215 (39%)	16 / 51 (31%)
<b>Circonstance de découverte – 3 catégories</b>				
DO	126 / 538 (23%)	0 / 10 (0%)	23 / 215 (11%)	1 / 51 (2.0%)
Médecin	312 / 538 (58%)	6 / 10 (60%)	106 / 215 (49%)	34 / 51 (67%)
Patient	100 / 538 (19%)	4 / 10 (40%)	86 / 215 (40%)	16 / 51 (31%)
<b>Statut hormonal</b>				
Basal like	47 / 538 (8.7%)	0 / 10 (0%)	19 / 215 (8.8%)	5 / 51 (9.8%)
Her2 surexprimé	27 / 538 (5.0%)	2 / 10 (20%)	16 / 215 (7.4%)	3 / 51 (5.9%)
Luminal	464 / 538 (86%)	8 / 10 (80%)	180 / 215 (84%)	43 / 51 (84%)
<b>FDep (quintiles)</b>				
1 (Très favorisé)	36 / 538 (6.7%)	0 / 10 (0%)	16 / 215 (7.4%)	4 / 51 (7.8%)
2	212 / 538 (39%)	5 / 10 (50%)	76 / 215 (35%)	13 / 51 (25%)
3	124 / 538 (23%)	0 / 10 (0%)	43 / 215 (20%)	20 / 51 (39%)
4	104 / 538 (19%)	3 / 10 (30%)	49 / 215 (23%)	12 / 51 (24%)
5 (Très défavorisé)	62 / 538 (12%)	2 / 10 (20%)	31 / 215 (14%)	2 / 51 (3.9%)

Caractéristiques	Local limite N = 538 <sup>1</sup>	Local étendu N = 10 <sup>1</sup>	Régional N = 215 <sup>1</sup>	Avance N = 51 <sup>1</sup>
<b>APL</b>				
Inférieur à la moyenne	147 / 538 (27%)	4 / 10 (40%)	46 / 215 (21%)	14 / 51 (27%)
Supérieur à la moyenne	391 / 538 (73%)	6 / 10 (60%)	169 / 215 (79%)	37 / 51 (73%)
<b>Baccalauréat</b>				
Non	168 / 538 (31%)	5 / 10 (50%)	75 / 215 (35%)	17 / 51 (33%)
Oui	370 / 538 (69%)	5 / 10 (50%)	140 / 215 (65%)	34 / 51 (67%)
<b>Temps</b>				
Commune avec Mammo	237 / 538 (44%)	4 / 10 (40%)	103 / 215 (48%)	23 / 51 (45%)
Moins de 15min	90 / 538 (17%)	1 / 10 (10%)	35 / 215 (16%)	5 / 51 (9.8%)
Entre 15 et 30min	108 / 538 (20%)	1 / 10 (10%)	42 / 215 (20%)	13 / 51 (25%)
Plus de 30min	103 / 538 (19%)	4 / 10 (40%)	35 / 215 (16%)	10 / 51 (20%)

Tableau 6 : Caractéristiques cliniques, sociodémographiques et diagnostiques des patientes atteintes de cancer du sein selon le stade d'extension au diagnostic, jeu de données « cas complet »

**Annexe 12. Stade au diagnostic en deux catégories, jeu de données avant imputation**

Caractéristiques	limité N = 597 <sup>1</sup>	évolué N = 287 <sup>1</sup>	p-value <sup>2</sup>
<b>Morphologie</b>			<b>&gt;0.9</b>
Carcinome canalaire	449 / 597 (75%)	214 / 287 (75%)	
Carcinome lobulaire	92 / 597 (15%)	46 / 287 (16%)	
Carcinome SAI et autres	56 / 597 (9.4%)	27 / 287 (9.4%)	
<b>Classe d'âge</b>			<b>0.033</b>
[15-40[	17 / 597 (2.8%)	16 / 287 (5.6%)	
[40-50[	70 / 597 (12%)	46 / 287 (16%)	
[50-75[	361 / 597 (60%)	151 / 287 (53%)	
75 et +	149 / 597 (25%)	74 / 287 (26%)	
<b>Année de diagnostic</b>			<b>0.2</b>
2017	209 / 597 (35%)	86 / 287 (30%)	
2018	196 / 597 (33%)	93 / 287 (32%)	
2019	192 / 597 (32%)	108 / 287 (38%)	
<b>Circonstance de découverte – 3 catégories</b>			<b>&lt;0.001</b>
DO	126 / 548 (23%)	24 / 266 (9.0%)	
Médecin	318 / 548 (58%)	140 / 266 (53%)	
Patient	104 / 548 (19%)	102 / 266 (38%)	
<b>Statut hormonal</b>			<b>0.4</b>
Basal like	48 / 597 (8.0%)	25 / 287 (8.7%)	
Her2 surexprimé	32 / 597 (5.4%)	22 / 287 (7.7%)	
Luminal	517 / 597 (87%)	240 / 287 (84%)	
<b>Département de résidence</b>			<b>0.002</b>
Corse-du-Sud	263 / 597 (44%)	158 / 287 (55%)	
Haute-Corse	334 / 597 (56%)	129 / 287 (45%)	

Caractéristiques	limité N = 597 <sup>1</sup>	évolué N = 287 <sup>1</sup>	p-value <sup>2</sup>
<b>FDep (quintiles)</b>			<b>0.8</b>
1 (Très favorisé)	41 / 597 (6.9%)	22 / 287 (7.7%)	
2	228 / 597 (38%)	98 / 287 (34%)	
3	140 / 597 (23%)	71 / 287 (25%)	
4	120 / 597 (20%)	63 / 287 (22%)	
5 (Très défavorisé)	68 / 597 (11%)	33 / 287 (11%)	
<b>APL</b>			<b>0.3</b>
Inférieur à la moyenne	162 / 597 (27%)	69 / 287 (24%)	
Supérieur à la moyenne	435 / 597 (73%)	218 / 287 (76%)	
<b>Baccalauréat</b>			<b>0.9</b>
Non	195 / 597 (33%)	95 / 287 (33%)	
Oui	402 / 597 (67%)	192 / 287 (67%)	
<b>Temps</b>			<b>0.5</b>
Commune avec Mammo	258 / 597 (43%)	136 / 287 (47%)	
Moins de 15min	102 / 597 (17%)	43 / 287 (15%)	
Entre 15 et 30min	120 / 597 (20%)	60 / 287 (21%)	
Plus de 30min	117 / 597 (20%)	48 / 287 (17%)	

Tableau 7 : Caractéristiques cliniques, sociodémographiques et diagnostiques des patientes atteintes de cancer du sein selon le stade d'extension au diagnostic en deux catégories et p-value du test du Chi<sup>2</sup> de Pearson

**Annexe 13. Hauteurs de coupes et dendrogramme du regroupement communal en profils communaux**

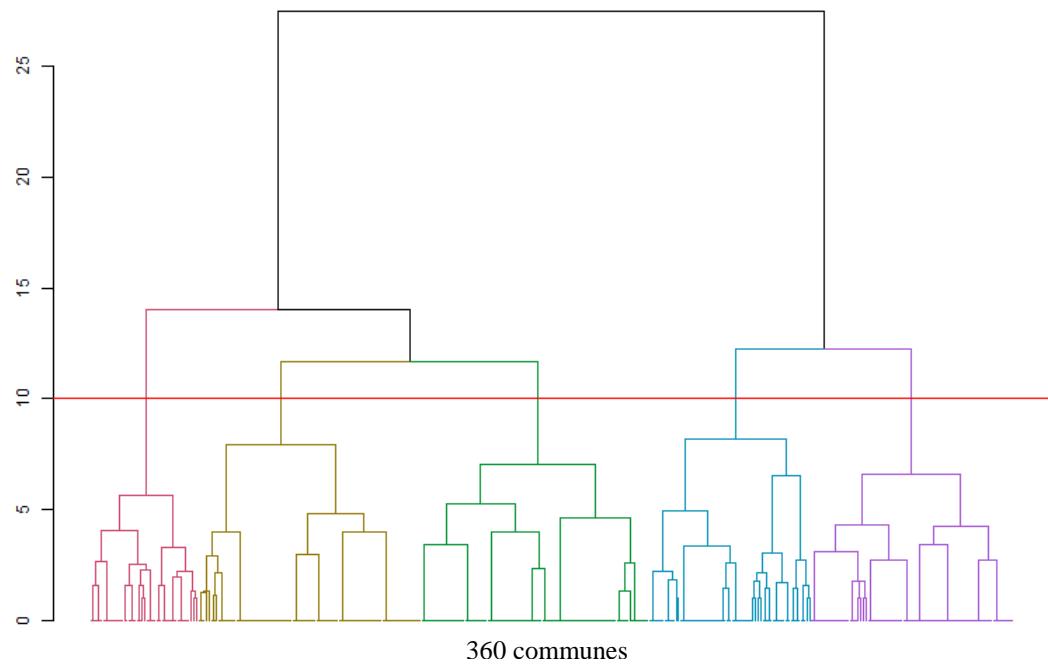


Figure 15 : Dendrogramme du regroupement communal en profils communaux en cinq classes

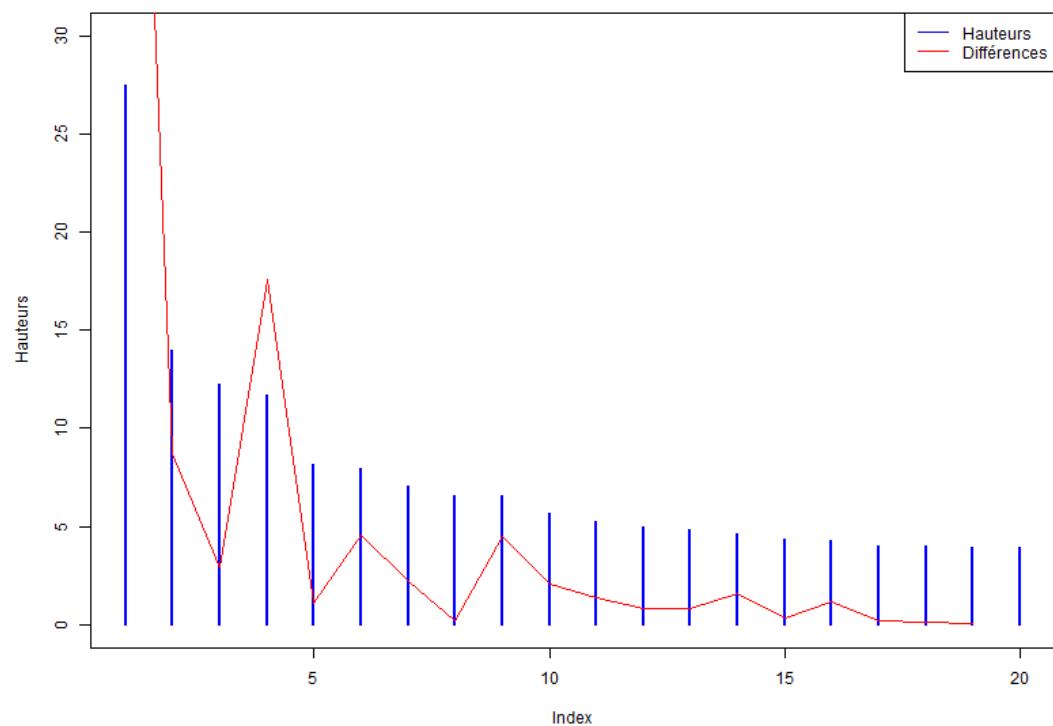


Figure 16 : Hauteurs de coupe et différence de hauteur de coupe en fonction du nombre de classes créées

**Annexe 14. Répartition des variables sociodémographiques en fonction de la classe communale d'appartenance**

Caractéristiques	Urbain, favorisé N = 42	Rural, favorisé N = 87	Urbain, défavorisé N = 63	Rural, interm. N = 89	Rural, défavorisé N = 79
<b>FDep (quintiles)</b>					
1 (Très favorisé)	13 (31%)	29 (33%)	0 (0%)	0 (0%)	0 (0%)
2	25 (60%)	58 (67%)	0 (0%)	12 (13%)	0 (0%)
3	4 (9.5%)	0 (0%)	11 (17%)	77 (87%)	0 (0%)
4	0 (0%)	0 (0%)	44 (70%)	0 (0%)	38 (48%)
5 (Très défavorisé)	0 (0%)	0 (0%)	8 (13%)	0 (0%)	41 (52%)
<b>APL</b>					
Inférieur à la moyenne	25 (60%)	74 (85%)	33 (52%)	49 (55%)	58 (73%)
Supérieur à la moyenne	17 (40%)	13 (15%)	30 (48%)	40 (45%)	21 (27%)
<b>Temps</b>					
Commune avec Mammo	2 (4.8%)	0 (0%)	6 (9.5%)	0 (0%)	0 (0%)
Moins de 15min	23 (55%)	0 (0%)	8 (13%)	0 (0%)	0 (0%)
Entre 15 et 30min	17 (40%)	18 (21%)	45 (71%)	21 (24%)	8 (10%)
Plus de 30min	0 (0%)	69 (79%)	4 (6.3%)	68 (76%)	71 (90%)
<b>Type</b>					
Bourg	14 (33%)	3 (3.4%)	23 (37%)	0 (0%)	0 (0%)
Rural_dispersé	27 (64%)	33 (38%)	28 (44%)	53 (60%)	40 (51%)
Rural_très_dispersé	1 (2.4%)	51 (59%)	12 (19%)	36 (40%)	39 (49%)

Tableau 8 : Caractéristiques sociodémographiques de classes communales créées

**Annexe 15.****Liste des communes selon la classe communale**

<b>Classe communale</b>	<b>Libellé des communes</b>
Urbain, favorisé	AFA, AJACCIO, ALATA, APPIETTO, AREGNO, BASTELICACCIA, BIGUGLIA, BRANDO, CASANOVA, CAURO, CORBARA, CORTE, COSTA, CUTTOLI CORTICCHIATO, ECCICA SUARELLA, ERBAJOLO, FELICETO, FURIANI, GRANACE, MONTICELLO, NESSA, OLMETO, ORTALE, PIETROSELLA, PIGNA, POGGIO D OLETTA, POGGIO DI VENACO, RIVENTOSA, SAN MARTINO DI LOTA, SANTA MARIA DI LOTA, SANTA REPARATA DI BALAGNA, SANTO PIETRO DI VENACO, SARROLA CARCOPINO, SERMANO, SOVERIA, VALLE DI MEZZANA, VENACO, VENTISERI, VERO, VIGGIANELLO, VILLANOVA, VILLE DI PIETRABUGNO
Rural, favorisé	AITI, ALANDO, ALBITRECCIA, ALTIANI, AMPRIANI, ARRO, ASCO, BARRETTALI, BELVEDERE CAMPOMORO, BILIA, BISINCHI, CAMBIA, CAMPANA, CAMPI, CAMPILE, CANAVAGGIA, CARCHETO BRUSTICO, CARGESE, CARTICASI, CASTIFAO, CHISA, COGGIA, COTI CHIAVARI, CRISTINACCE, ERONE, FARINOLE, FAVALELLO, FELCE, FOCE, FOCICCHIA, FORCIOLO, FRASSETTO, GALERIA, GAVIGNANO, GROSSA, GROSSETO PRUGNA, GUAGNO, GUITERA LES BAINS, LAMA, LETIA, LOZZI, MARIGNANA, MATRA, MAUSOLEO, MAZZOLA, MONACIA D AULLENE, MONACIA D OREZZA, MURACCIOLE, MURZO, NOCETA, OLCANI, OLIVESE, PARATA, PERI, PIANO, PIANOTTOLI CALDARELLO, PIEDICROCE, PIETRASERENA, PIETRICAGGIO, PIOBETTA, PIOGGIOLA, POLVEROSO, RAPAGGIO, RAPALE, REZZA, ROSAZIA, RUSIO, SALICETO, SAN GAVINO DI TENDA, SAN LORENZO, SANT ANDREA DI BOZIO, SANTA LUCIA DI MERCURIO, SANTA MARIA SICHE, SCATA, SERRIERA, SOCCIA, SORBOLANO, SORIO, STAZZONA, TARRANO, TOLLA, TOX, UCCIANI, VALLE DI CAMPOLORO, ZIGLIARA, ZOZA, ZUANI
Urbain, défavorisé	ALERIA, ALGAJOLA, ALTAGENE, ARBELLARA, BASTIA, BONIFACIO, BORGO, CALACUCCIA, CALVI, CAMPITELLO, CARGIACA, CASALABRIVA, CASTELLARE DI CASINCA, CASTELLARE DI MERCURIO, CASTIRLA, CATERI, CONCA, FIGARI, FOZZANO, GHISONACCIA, GHISONI, L ILE ROUSSE, LAVATOGGIO, LORETO DI TALLANO, LUCCIANA, LUGO DI NAZZA, MONTE, MONTEGROSSO, MOROSAGLIA, MURATO, MURO, OCCHIATANA, OLMETA DI TUDA, OLMICCIA, OMessa, PALASCA, PATRIMONIO, PENTA DI CASINCA, PIETRACORBARA, PIETROSO, POGGIO DI NAZZA, PORTO VECCHIO, PRATO DI GIOVELLINA, PROPRIANO, PRUNELLI DI CASACCONI, PRUNELLI DI FIUMORBO, PRUNO, SAN GAVINO DI FIUMORBO, SAN NICOLAO, SANTA LUCIA DI MORIANI, SARTENE, SERRA DI FIUMORBO, SOLLACARO, SORBO OCAGNANO, ST FLORENT, STE LUCIE DE TALLANO, TRALONCA, VALLECALLE, VENZOLASCA, VESCOVATO, VILLE DI PARASO, VOLPAJOLA, ZONZA

Classe communale	Libellé des communes
Rural, intermédiaire	ALBERTACCE, AMBIEGNA, ARBORI, AVAPESSA, AZILONE AMPAZA, BALOGNA, BARBAGGIO, BASTELICA, BELGODERE, BOCOGNANO, CALCATOGGIO, CAMPO, CANNELLE, CARBINI, CARDÒ TORGIA, CASABIANCA, CASAGLIONE, CASAMACCIOLI, CASTELLO DI ROSTINO, CASTINETA, CERVIONE, CHIATRA, CIAMANNACCE, CORRANO, CORSIA, Ersa, EVISA, FICAJA, GIOCATOJO, GIUNCAGGIO, GIUNCHETO, GUARGUALE, LECCI, LENTO, LEVIE, LOPIGNA, LORETO DI CASINCA, LUMIO, MANSO, MOCA CROCE, MOLTIFAO, NOVALE, NOVELLA, OCANA, OLETTA, ORTO, PENTA ACQUATELLA, PERELLI, PETRETO BICCHISANO, PIANELLO, PIAZZOLE, PIEDICORTE DI GAGGIO, PIEDIGRIGGIO, PIEDIPARTINO, PIETRALBA, PIEVE, POGGIO MARINACCIO, POGGIOLO, QUASQUARA, RUTALI, SALICE, SAN GAVINO DI CARBINI, SAN GIOVANNI DI MORIANI, SANT ANDREA D ORCINO, SANT ANDREA DI COTONE, SANT ANTONINO, SANTO PIETRO DI TENDA, SARI D ORCINO, SCOLCA, SERRA DI FERRO, SERRA DI SCOPAMENE, SILVARECCIO, SISCO, SOLARO, SOTTA, SPELONCATO, TAGLIO ISOLACCIO, TAVACO, TAVERA, URBALACONE, URTACA, VALLE D ALESANI, VALLE D OREZZA, VALLE DI ROSTINO, VICO, VIVARIO, ZERUBIA, ZEVACO, ZICAVO
Rural, défavorisé	AGHIONE, ALZI, ANTISANTI, ARGIUSTA MORICCIO, AULENE, AZZANA, BIGORNO, BUSTANICO, CAGNANO, CALENZANA, CANALE DI VERDE, CANARI, CARBUCCIA, CARPINETO, CASALTA, CASEVECCHIE, CASTIGLIONE, CENTURI, COGNOCOLI MONTICCHI, COZZANO, CROCE, CROCICCHIA, ISOLACCIO DI FIUMORBO, LA PORTA, LANO, LINGUIZZETTA, LURI, MELA, MERIA, MOITA, MONCALE, MORSIGLIA, NOCARIO, NONZA, OGLIASTRO, OLMETA DI CAPOCORSO, OLMI CAPPELLA, OLMO, ORTIPORIO, OSANI, OTA, PALNECA, PANCHERACCIA, PARTINELLO, PASTRICCIOLA, PERO CASEVECCHIE, PIANA, PIAZZALI, PIE D OREZZA, PIETRA DI VERDE, PILA CANALE, PINO, POGGIO MEZZANA, POPOLASCA, PORRI, QUENZA, QUERCITELLO, RENNO, ROGLIANO, ROSPIGLIANI, SAMPOLO, SAN DAMIANO, SAN GAVINO D AMPUGNANI, SAN GIULIANO, SANTA MARIA FIGIANELLA, SANTA MARIA POGGIO, SANTA REPARATA DI MORIANI, SARI SOLENZARA, TALASANI, TALLONE, TASSO, TOMINO, VALICA, VELONE ORNETO, VERDESE, VEZZANI, VIGNALE, ZALANA, ZILIA

Tableau 9 : Liste des libellés des communes selon les cinq classes communales établies

## Annexe 16. Matrices de Corrélation

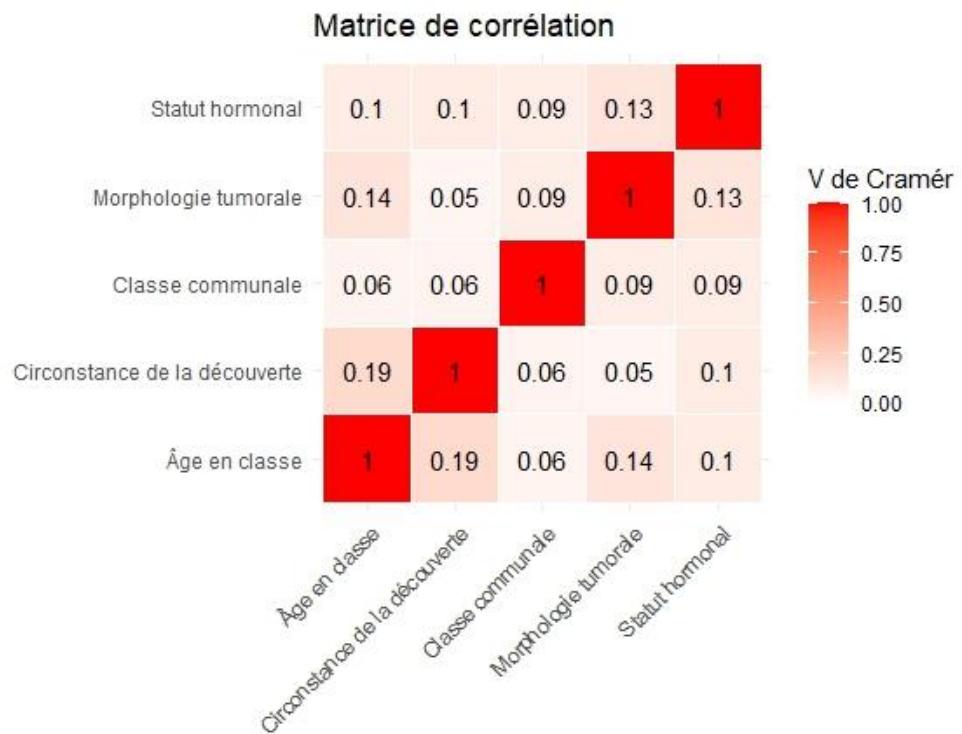


Figure 17 : Matrice de corrélation calculée pour les variables à inclure dans le modèle complet, étude en cas complet

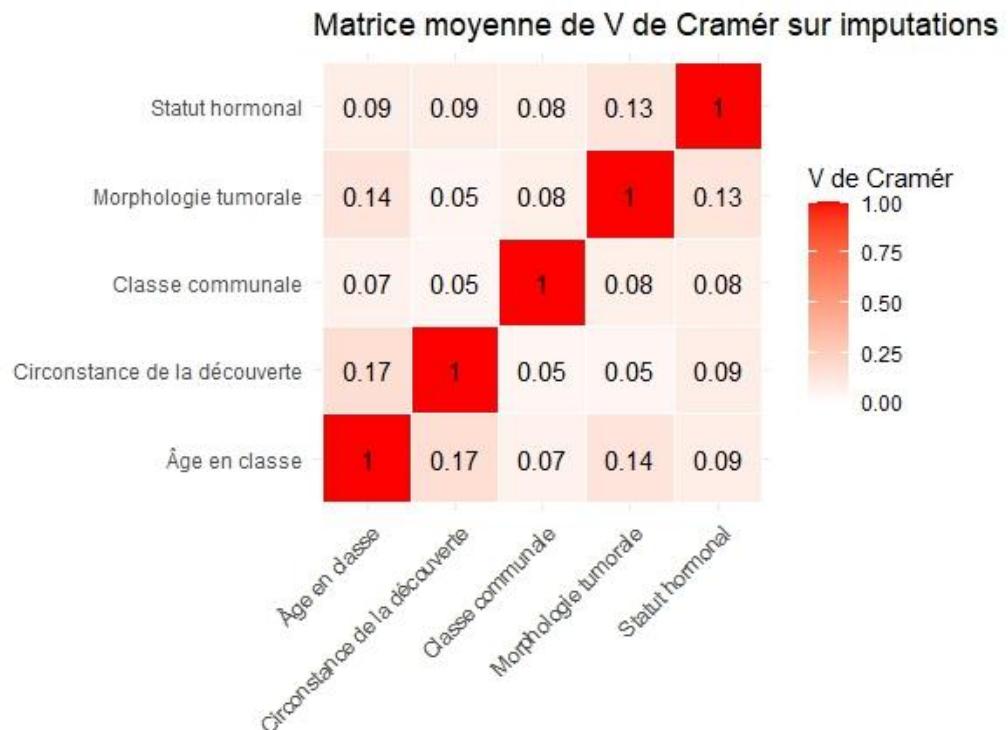


Figure 18 : Moyenne des matrices de corrélation calculée, pour chaque imputation, pour les variables à inclure dans le modèle complet, données imputées

## Annexe 17. Analyse de la déviance, cas complet

L'analyse de la déviance consiste en une comparaison entre le modèle complet ci-dessous et un modèle emboité dans lequel une variable a été retirée.

Modèle complet :

$$\begin{aligned}
 & \text{logit}(P(\text{Stade au diagnostic} = \text{élevé})) \\
 & = \beta_0 + \beta_1 \text{Morphologie tumorale} + \beta_2 \text{Classe d'âge} \\
 & + \beta_3 \text{Circonstance de la découverte} + \beta_4 \text{Statut hormonal} + \beta_5 \text{Classe communale}
 \end{aligned}$$

Variable retirée	Valeur p associé au test du rapport de vraisemblance	Résultat du test
Classe communale	0,2969	Non significatif
Morphologie tumorale	0,7396	Non significatif
Classe d'âge	0,302	Non significatif
Circonstances de la découverte	<<0,01	Significatif
Statut hormonal	0,996	Non significatif

Tableau 10 : Analyse de l'apport individuel des variables présentes dans le modèle logistique, analyse en cas complet

**Annexe 18. Exponentielle des coefficients et leurs intervalles de confiance, cas complet**

Variable : modalité	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0,20	0,52	-3,10	0,002	0,07	0,54
Morphologie : Carcinome lobulaire	0,84	0,22	-0,77	0,443	0,54	1,29
Morphologie : Autres carcinomes	0,99	0,27	-0,03	0,978	0,58	1,67
Classe d'âge : [40-50[ ans	1,22	0,42	0,48	0,633	0,53	2,84
Classe d'âge : [50-75[ ans	0,82	0,39	-0,52	0,601	0,38	1,77
Classe d'âge : 75 ans et +	0,78	0,40	-0,61	0,542	0,35	1,74
Circonstance : Medecin	2,25	0,25	3,26	0,001	1,40	3,73
Circonstance : Patient	5,01	0,27	5,90	0,000	2,97	8,69
Statut hormonal : Her2+	1,03	0,41	0,07	0,942	0,46	2,28
Statut hormonal :Luminal	1,02	0,28	0,09	0,932	0,60	1,79
Commune : Rural, favorisé	0,98	0,34	-0,07	0,944	0,49	1,88
Commune : Urbain, défavorisé	1,30	0,18	1,48	0,139	0,92	1,83
Commune : Rural, intermédiaire	1,68	0,30	1,74	0,082	0,93	3,01
Commune : Rural, défavorisé	0,92	0,34	-0,25	0,801	0,46	1,77

Tableau 11 : Exponentielle des coefficients, écarts-type, statistique de test, p-value et bornes de l'intervalle de confiance à 95% des modalités du modèle de régression logistique de l'analyse en cas complet

## Annexe 19. Analyse de la déviance, données imputées

L'analyse de la déviance consiste en une comparaison entre le modèle complet ci-dessous et un modèle emboité dans lequel une variable a été retirée.

Modèle complet :

$$\begin{aligned} & \text{logit}(P(\text{Stade au diagnostic} = \text{élevé})) \\ & = \beta_0 + \beta_1 \text{Morphologie tumorale} + \beta_2 \text{Classe d'âge} \\ & + \beta_3 \text{Circonstance de la découverte} + \beta_4 \text{Statut hormonal} + \beta_5 \text{Classe communale} \end{aligned}$$

Variable retirée	Valeur p associé au test D3	Résultat du test
Classe communale	0,407	Non significatif
Morphologie tumorale	0,956	Non significatif
Classe d'âge	0,0,506	Non significatif
Circonstances de la découverte	<<0,01	Significatif
Statut hormonal	0,846	Non significatif

Tableau 12 : Analyse de l'apport individuel via un test D3 des variables présentes dans le modèle logistique, analyse avec des données imputées

## Annexe 20. Exponentielle des coefficients et leurs intervalles de confiance, données imputées

Variable : modalité	estimate	std.error	statistic	p.value	conf.low	conf.high
(Intercept)	0,27	0,51	-2,52	0,012	0,10	0,75
Morphologie : Carcinome lobulaire	1,06	0,21	0,26	0,792	0,70	1,59
Morphologie : Autres carcinomes	0,98	0,26	-0,09	0,926	0,58	1,63
Classe d'âge : [40-50[ ans	0,95	0,41	-0,13	0,896	0,42	2,13
Classe d'âge : [50-75[ ans	0,71	0,38	-0,89	0,371	0,34	1,50
Classe d'âge : 75 ans et +	0,72	0,39	-0,85	0,395	0,33	1,55
Circonstance : Medecin	2,06	0,25	2,92	0,004	1,27	3,34
Circonstance : Patient	4,18	0,28	5,19	0,000	2,43	7,18
Statut hormonal : Her2+	1,10	0,39	0,25	0,802	0,52	2,35
Statut hormonal : Luminal	0,93	0,27	-0,25	0,800	0,55	1,59
Commune : Rural, favorisé	0,86	0,34	-0,45	0,651	0,45	1,66
Commune : Urbain, défavorisé	1,16	0,17	0,91	0,363	0,84	1,62
Commune : Rural, intermédiaire	1,57	0,28	1,59	0,111	0,90	2,72
Commune : Rural, défavorisé	0,86	0,34	-0,45	0,650	0,44	1,67

Figure 19 : Exponentielle des coefficients, écarts-type, statistique de test, p-value et bornes de l'intervalle de confiance à 95% des modalités du modèle de régression logistique de l'analyse contenant des données imputées

**Annexe 21. Caractéristiques des patientes selon les circonstances de la découverte détaillées**

Caractéristiques	Auto-palpation N = 148	Rétraction N = 45	Examen cli. N = 117	DI N = 289	Bilan N = 54	DO N = 150
<b>Classe d'âge</b>						
[15-40[	13 / 148 (8.8%)	3 / 45 (6.7%)	5 / 117 (4.3%)	8 / 289 (2.8%)	1 / 54 (1.9%)	0 / 150 (0%)
[40-50[	26 / 148 (18%)	5 / 45 (11%)	16 / 117 (14%)	35 / 289 (12%)	10 / 54 (19%)	9 / 150 (6.0%)
[50-75[	65 / 148 (44%)	20 / 45 (44%)	56 / 117 (48%)	185 / 289 (64%)	25 / 54 (46%)	118 / 150 (79%)
75 et +	44 / 148 (30%)	17 / 45 (38%)	40 / 117 (34%)	61 / 289 (21%)	18 / 54 (33%)	23 / 150 (15%)
<b>stade ENCR</b>						
Local limite	74 / 148 (50%)	17 / 45 (38%)	73 / 117 (62%)	208 / 289 (72%)	33 / 54 (61%)	126 / 150 (84%)
Local etendu	2 / 148 (1.4%)	2 / 45 (4.4%)	6 / 117 (5.1%)	0 / 289 (0%)	0 / 54 (0%)	0 / 150 (0%)
Regional	66 / 148 (45%)	17 / 45 (38%)	29 / 117 (25%)	68 / 289 (24%)	9 / 54 (17%)	23 / 150 (15%)
Avance	6 / 148 (4.1%)	9 / 45 (20%)	9 / 117 (7.7%)	13 / 289 (4.5%)	12 / 54 (22%)	1 / 150 (0.7%)
<b>Stade 2 cat.</b>						
limité	76 / 148 (51%)	19 / 45 (42%)	79 / 117 (68%)	208 / 289 (72%)	33 / 54 (61%)	126 / 150 (84%)
étendu	72 / 148 (49%)	26 / 45 (58%)	38 / 117 (32%)	81 / 289 (28%)	21 / 54 (39%)	24 / 150 (16%)
<b>Classe commune</b>						
Urbain, favorisé	62 / 148 (42%)	13 / 45 (29%)	36 / 117 (31%)	117 / 289 (40%)	17 / 54 (31%)	59 / 150 (39%)
Rural, favorisé	11 / 148 (7.4%)	3 / 45 (6.7%)	9 / 117 (7.7%)	17 / 289 (5.9%)	2 / 54 (3.7%)	9 / 150 (6.0%)
Urbain, défavorisé	52 / 148 (35%)	17 / 45 (38%)	55 / 117 (47%)	118 / 289 (41%)	31 / 54 (57%)	62 / 150 (42%)
Rural, intermédiaire	13 / 148 (8.8%)	6 / 45 (13%)	8 / 117 (6.8%)	23 / 289 (8.0%)	2 / 54 (3.7%)	10 / 150 (6.6%)
Rural, défavorisé	10 / 148 (6.8%)	6 / 45 (13%)	9 / 117 (7.7%)	14 / 289 (4.8%)	2 / 54 (3.7%)	10 / 150 (6.6%)

Tableau 13 : Caractéristiques clinique et sociodémographique des patientes selon les circonstances détaillées de la découverte

## Annexe 22.

## Analyses de la survie

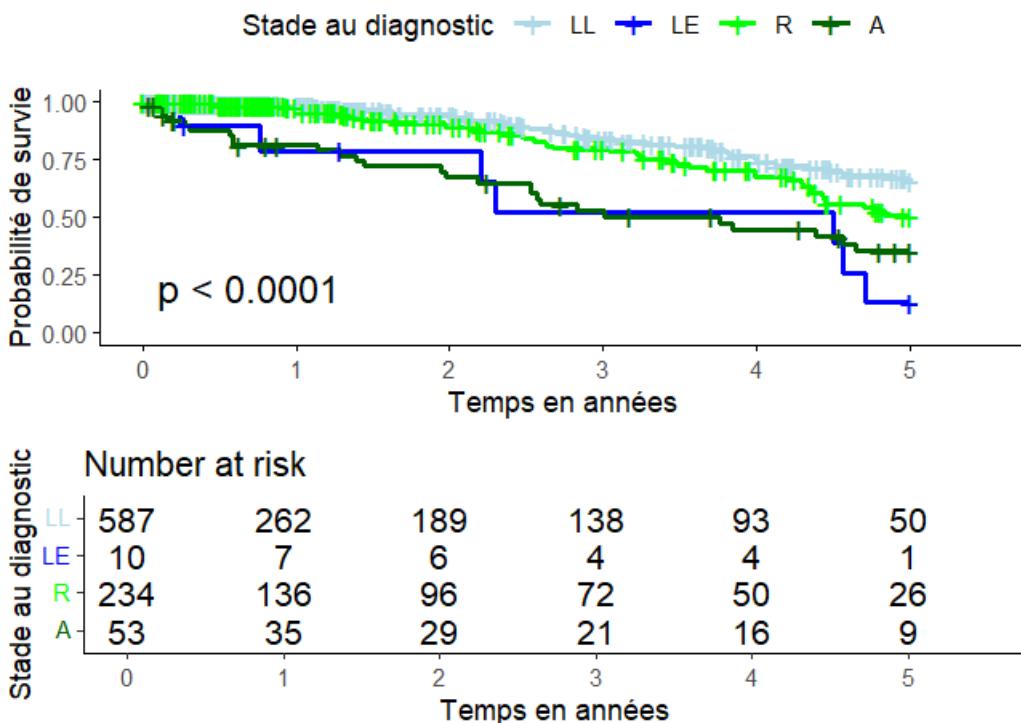


Figure 20 : Courbes de survie globale à cinq ans calculées en fonction du stade au diagnostic en quatre catégories avec la méthode de Kaplan-Meier, valeur de la  $p$ -value du test du log-rank et tableau du nombre d'individus à risque aux différents temps.

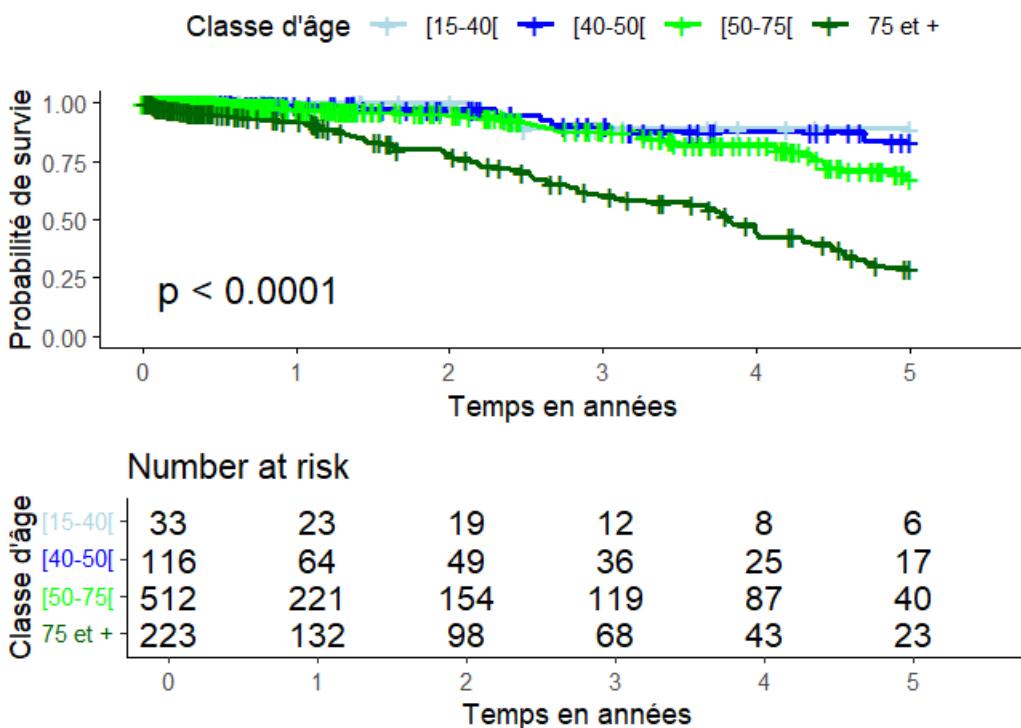


Figure 21 : Courbes de survie globale à cinq ans calculées en fonction de l'âge en quatre catégories avec la méthode de Kaplan-Meier, valeur de la  $p$ -value du test du log-rank et tableau du nombre d'individus à risque aux différents temps.

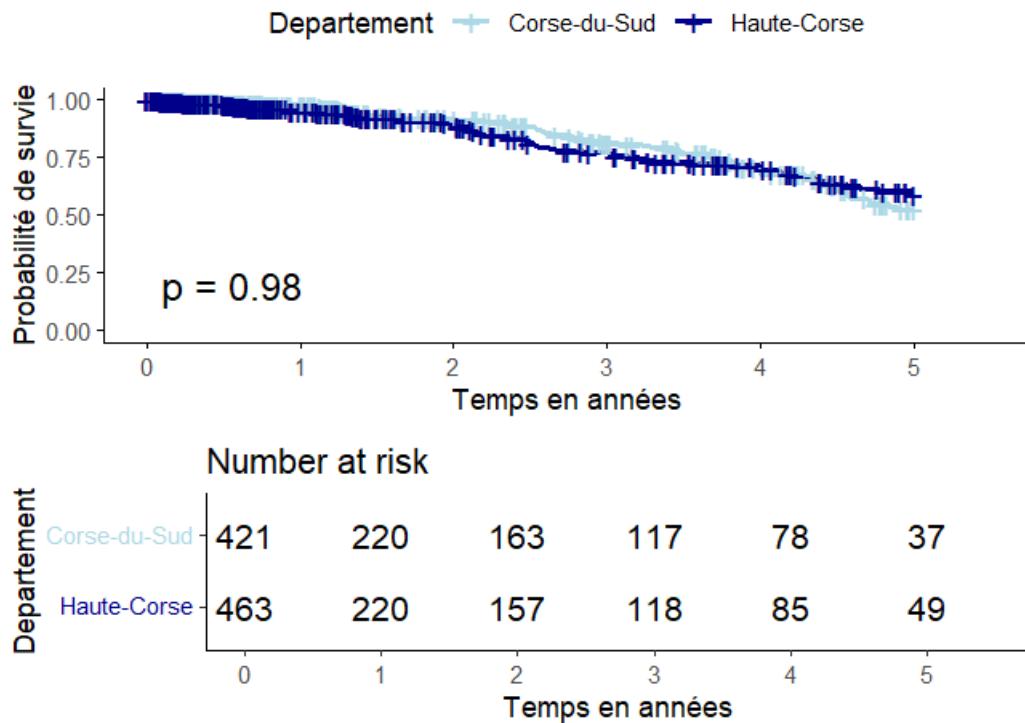


Figure 22 : Courbes de survie globale à cinq ans calculées en fonction du département de résidence avec la méthode de Kaplan-Meier, valeur de la  $p$ -value du test du log-rank et tableau du nombre d'individus à risque aux différents temps. I

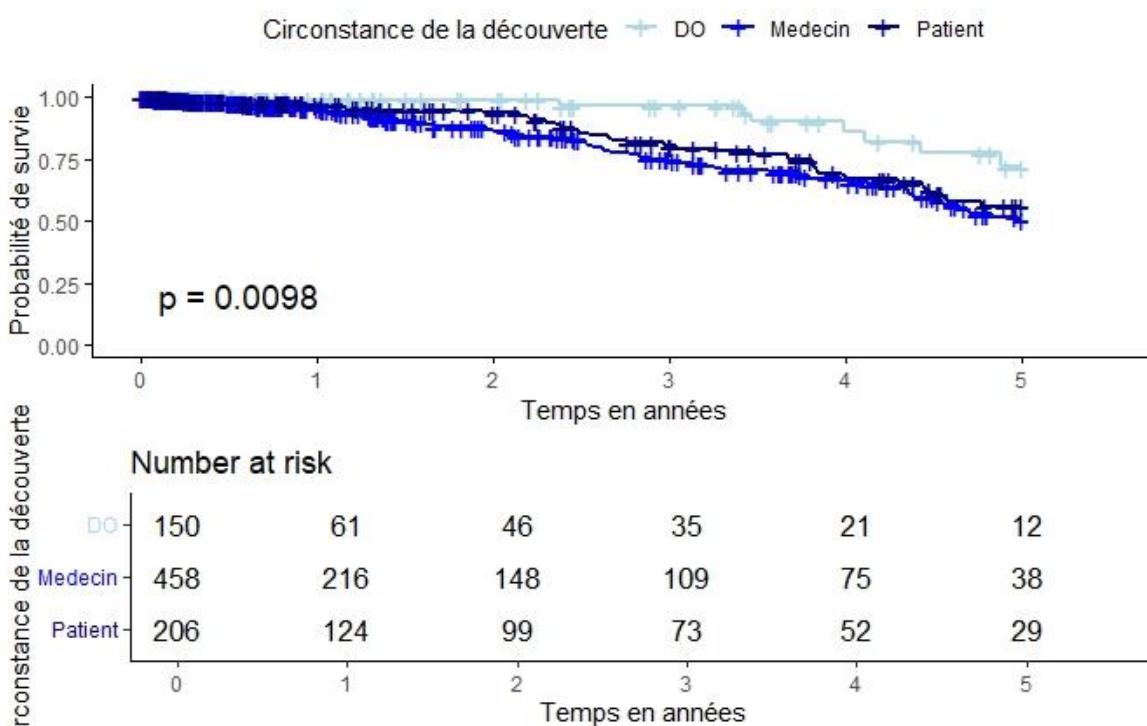


Figure 23 : Courbes de survie globale à cinq ans calculées en fonction des circonstances de la découverte avec la méthode de Kaplan-Meier, valeur de la  $p$ -value du test du log-rank et tableau du nombre d'individus à risque aux différents temps.

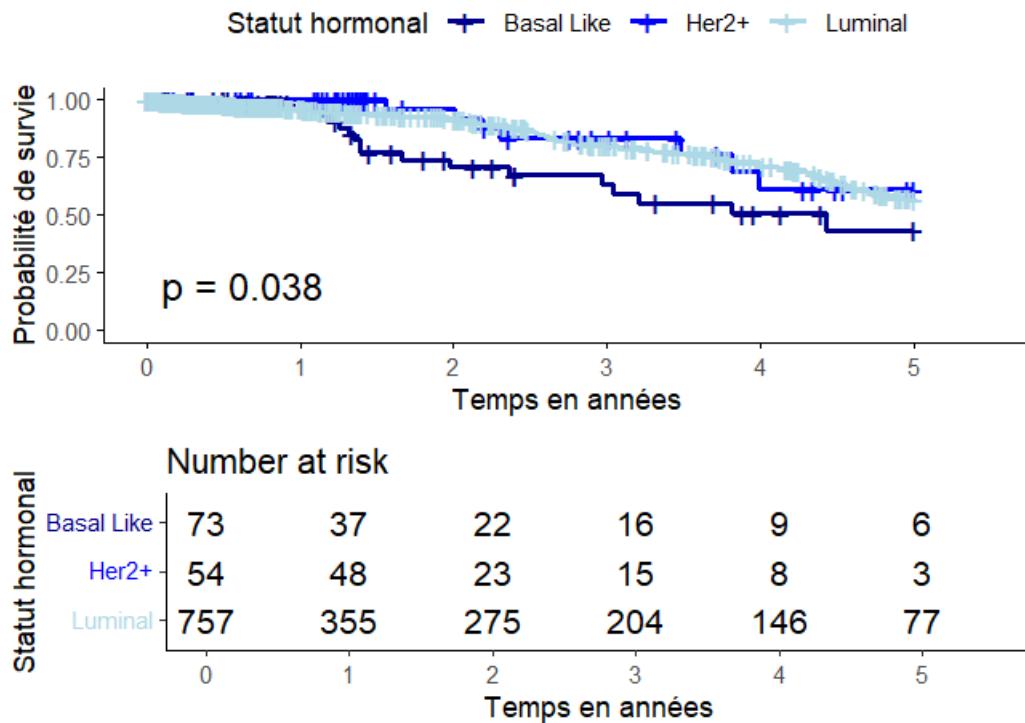


Figure 24 : Courbes de survie globale à cinq ans calculées en fonction du statut hormonal de la tumeur avec la méthode de Kaplan-Meier, valeur de la p-value du test du log-rank et tableau du nombre d'individus à risque aux différents temps.

## Références

1. Heuzé G, Cariou M, Billot-Grasset A, Chatignoux E. Estimations régionales et départementales d'incidence et de mortalité par cancers en France, 2007 - 2016.
2. Arena C. Incidences 2017-2019, Méthodes – Résultats – Perspectives.
3. SPF. Stade au diagnostic des cancers du sein, du côlon et du rectum : Étude réalisée à partir des registres des cancers du réseau Francim [Internet]. 2018 [cited 2025 Apr 8]. Available from: <https://www.santepubliquefrance.fr/import/stade-au-diagnostic-des-cancers-du-sein-du-colon-et-du-rectum-etude-realisee-a-partir-des-registres-des-cancers-du-reseau-francim>
4. Aloulou S, El Mahfoudi A, El Omrani A, Khouchani M. Facteurs liés au diagnostic tardif du cancer du sein: expérience du CHU Mohammed VI Marrakech. Pan Afr Med J. 2015 Jun 25;21:162.
5. Berger F, Doussau A, Gautier C, Gros F, Asselain B, Reyal F. Impact du statut socioéconomique sur la gravité du diagnostic initial de cancer du sein. Revue d'Épidémiologie et de Santé Publique. 2012 Feb 1;60(1):19–29.
6. Delacote C, Ariza JM, Delacour-Billon S, Ayrault-Piault S, Borghi G, Menanteau K, Bouron A, Métais M, Cowppli-Bony A, Molinié F. Socioeconomic and geographic disparities of breast cancer incidence according to stage at diagnosis in France. Cancer Causes Control. 2024 Feb 1;35(2):241–51.
7. Borghi G, Delacote C, Delacour-Billon S, Ayrault-Piault S, Dabakuyo-Yonli TS, Delafosse P, Woronoff AS, Trétarre B, Molinié F, Cowppli-Bony A. Socioeconomic Deprivation and Invasive Breast Cancer Incidence by Stage at Diagnosis: A Possible Explanation to the Breast Cancer Social Paradox. Cancers. 2024 Jan;16(9):1701.
8. Rogel A. Couverture totale du dépistage du cancer du sein, dans et hors programme organisé, de 2016 à 2022 en France hexagonale / Breast cancer screening coverage within and outside the organised programme, 2016–2022, mainland France.
9. Taux de participation au programme de dépistage organisé du cancer du sein 2021-2022 et évolution depuis 2005 [Internet]. [cited 2025 Apr 8]. Available from: <https://www.santepubliquefrance.fr/maladies-et-traumatismes/cancers/cancer-du-sein/articles/taux-de-participation-au-programme-de-depistage-organise-du-cancer-du-sein-2021-2022-et-evolution-depuis-2005>
10. Synthèse – Panorama de la pauvreté en Corse : une diversité de situations individuelles et territoriales | Insee [Internet]. [cited 2025 Apr 10]. Available from: <https://www.insee.fr/fr/statistiques/7677733?sommaire=7678482>
11. Delacote C, Ariza JM, Delacour-Billon S, Ayrault-Piault S, Borghi G, Menanteau K, Bouron A, Métais M, Cowppli-Bony A, Molinié F. Socioeconomic and geographic disparities of breast cancer incidence according to stage at diagnosis in France. Cancer Causes Control. 2024 Feb 1;35(2):241–51.
12. Brierley JD, Gospodarowicz MK, Wittekind C. TNM Classification des tumeurs malignes, 8ème édition.
13. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sabin LH, Parkin DM, Whelan SL, World Health Organization. Classification internationale des maladies pour l'oncologie. International classification of diseases for oncology [Internet]. 3rd ed. 2008 [cited 2025 Apr 17]; Available from: <https://iris.who.int/handle/10665/43859>
14. Makki J. Diversity of Breast Carcinoma: Histological Subtypes and Clinical Relevance. Clin Med Insights Pathol. 2015 Dec 21;8:23–31.

15. Haïfa B, El Mhabrech H, Zouari I, Njima M, Daldoul A, Ahmed H, Awatef H, Khchine W, Zaidi S, Faleh R. Le carcinome lobulaire infiltrant du sein: à propos de 30 cas. *Pan Afr Med J.* 2019 Oct 4;34:70.
16. Frikha N, Chlif M. Un aperçu des facteurs de risque du cancer du sein. *Bulletin de l'Académie Nationale de Médecine.* 2021 May;205(5):519–27.
17. Reeves GK, Beral V, Green J, Gathani T, Bull D. Hormonal therapy for menopause and breast-cancer risk by histological type: a cohort study and meta-analysis. *The Lancet Oncology.* 2006 Nov 1;7(11):910–8.
18. Li CI, Uribe DJ, Daling JR. Clinical characteristics of different histologic types of breast cancer. *Br J Cancer.* 2005 Oct 31;93(9):1046–52.
19. Rey G, Jouglé E, Fouillet A, Hémon D. 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.* 2009 Jan 22;9:33.
20. Giraud T. osrm: Interface Between R and the OpenStreetMap-BasedRouting Service OSRM. *JOSS.* 2022 Oct 27;7(78):4574.
21. Les hameaux : de « nouveaux villages » ou quand le développement démographique se fait hors du chef-lieu - Insee Analyses Corse - 13 [Internet]. [cited 2025 May 14]. Available from: <https://www.insee.fr/fr/statistiques/2590380>
22. Dahlman D, Magnusson H, Li X, Sundquist J, Sundquist K. Drug use disorder and risk of incident and fatal breast cancer: a nationwide epidemiological study. *Breast Cancer Res Treat.* 2021 Feb 1;186(1):199–207.
23. Barrault-Couchouron M, Béracochéa M, Dorval M, Allafort V, Barthélémy V, Garguil V, Auriacombe M, Bussières E, M'bailara K. Consommations de tabac et d'alcool chez des femmes traitées pour un cancer du sein dans un service de chirurgie oncologique : des comportements fréquents à considérer. *Bulletin du Cancer.* 2022 Mar 1;109(3):307–17.
24. Seppänen AV, Daniel F, Houzard S, Le Bihan C, Coldefy M, Gandré C. Breast Cancer Care Pathways for Women with Preexisting Severe Mental Disorders: Evidence of Disparities in France? *Journal of Clinical Medicine.* 2023 Jan;12(2):412.
25. Van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations [Internet]. 2006 [cited 2025 Feb 25]. p. 3.17.0. Available from: <https://CRAN.R-project.org/package=mice>
26. Graham JW, Olchowski AE, Gilreath TD. How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. *Prev Sci.* 2007 Sep 1;8(3):206–13.
27. Ward JH. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association.* 1963;58(301):236–44.
28. Meng XL, Rubin DB. Performing likelihood ratio tests with multiply-imputed data sets. *Biometrika.* 1992;79(1):103–11.
29. D3 function - RDocumentation [Internet]. [cited 2025 May 29]. Available from: <https://www.rdocumentation.org/packages/mice/versions/3.17.0/topics/D3>
30. Exarchakou A, Kipourou DK, Belot A, Rachet B. Socio-economic inequalities in cancer survival: how do they translate into Number of Life-Years Lost? *Br J Cancer.* 2022 Jun 1;126(10):1490–8.
31. Ribeiro AI, Fraga S, Severo M, Kelly-Irving M, Delpierre C, Stringhini S, Kivimaki M, Joost S, Guessous I, Severi G, Giles G, Sacerdote C, Vineis P, Barros H, Alberts J, Alenius H, Avendano M, Baglietto L, Baltar V, Bartley M, Barros H, Bellone M, Berger E, Blane D, Bochud M, Candiani G, Carmeli C, Carra L, Castagne R, Chadeau-Hyam M, Cima S, Costa G, Courtin E, Delpierre C, Donkin A, D'Errico A, Dugue PA, Elliot P, Fagherazzi

G, Fiorito G, Fraga S, Gandini M, Gares V, Gerbouin-Rerolle P, Giles G, Goldberg M, Greco D, Hodge A, Kelly-Irving M, Karimi M, Karisola P, Kivimaki M, Laine J, Lang T, Laurent A, Layte R, Lepage B, Lorsch D, Machell G, Mackenbach J, Mestral C de, McCrory C, Miller C, Milne R, Muennig P, Nusselder W, Petrovic D, Pilapil L, Polidoro S, Preisig M, Ribeiro AI, Ricceri F, Recalcati P, Reinhard E, Robinson O, Valverde JR, Saba S, Santegoets F, Simmons T, Severi G, Stringhini S, Tabak A, Terhi V, Tieulent J, Vaccarella S, Vignataglianti F, Vineis P, Vollenweider P, Zins M. Association of neighbourhood disadvantage and individual socioeconomic position with all-cause mortality: a longitudinal multicohort analysis. *The Lancet Public Health*. 2022 May 1;7(5):e447–57.

32. Lamy S, Molinié F, Daubisse-Marliac L, Cowppli-Bony A, Ayrault-Piault S, Fournier E, Woronoff AS, Delpierre C, Grosclaude P. Using ecological socioeconomic position (SEP) measures to deal with sample bias introduced by incomplete individual-level measures: inequalities in breast cancer stage at diagnosis as an example. *BMC Public Health*. 2019 Jul 2;19:857.
33. Haneuse SJPA, Wakefield JC. The Combination of Ecological and Case-Control Data. *J R Stat Soc Series B Stat Methodol*. 2008 Feb 1;70(1):73–93.
34. Plan\_cancer\_2009-2013.pdf [Internet]. [cited 2025 May 15]. Available from: [https://sante.gouv.fr/IMG/pdf/Plan\\_cancer\\_2009-2013.pdf](https://sante.gouv.fr/IMG/pdf/Plan_cancer_2009-2013.pdf)
35. Ebell MH, Thai TN, Royalty KJ. Cancer screening recommendations: an international comparison of high income countries. *Public Health Rev*. 2018 Mar 2;39:7.
36. Khrouf S, Ksontini FL, Ayadi M, Rais HBA, Mezlini A. Le dépistage du cancer du sein : une polémique qui divise. *Breast cancer screening: a dividing controversy*. LA TUNISIE MEDICALE. 2020;98.
37. Le dépistage organisé du cancer du sein [Internet]. [cited 2025 May 15]. Available from: <https://www.ameli.fr/medecin/sante-prevention/pathologies/cancers/depistage-organise-du-cancer-du-sein>
38. synthese\_et\_recommandations\_participation\_depistage\_cancer\_du\_sein.pdf [Internet]. [cited 2025 May 15]. Available from: [https://www.has-sante.fr/upload/docs/application/pdf/2012-07/synthese\\_et\\_recommandations\\_participation\\_depistage\\_cancer\\_du\\_sein.pdf](https://www.has-sante.fr/upload/docs/application/pdf/2012-07/synthese_et_recommandations_participation_depistage_cancer_du_sein.pdf)
39. Morel J. Évaluation des connaissances et de la pratique de l'autopalpation mammaire dans la population féminine française de 18 à 50 ans.
40. Buquet Z. La pratique de l'auto-examen des seins par les femmes françaises de 20 à 49 ans. Une enquête nationale observationnelle univariée et multivariée.
41. Lavoué V, Favier A, Frank S, Boutet G, Azuar AS, Brousse S, Golfier F, Uzan C, Vaysse C, Molière S, Boissière-Lacroix M, Kermarrec E, Seror JY, Delpech Y, Luporsi É, Maugard CM, Taris N, Chabbert-Buffet N, Sabah J, Alghamdi K, Fritel X, Mathelin C. Recommandations pour la pratique clinique du CNGOF. Place de l'auto-examen des seins dans les stratégies de dépistage. *Gynécologie Obstétrique Fertilité & Sénologie*. 2023 Oct;51(10):437–47.
42. Ab A. Medical Students' Knowledge and Attitude towards Breast Cancer Risk Factors and Early Detection Practices.